CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

Canada

NL 339 (r. 88/09)

20

0-315-01216-0

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Division
Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Jo-Anne LeFevre

Date of Birth — Date de naissance	Country of Birth — Lieu de naissance
August 5, 1960	Canada

Permanent Address — Résidence fixe

#305-10620-79 Avenue
Edmonton, Alberta
T6E 1S1

Title of Thesis — Titre de la thèse

A Model of the Use of Instruction and Example Information on a Simple Inductive-Reasoning Task

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

MSc

Year this degree conferred — Année d'obtention de ce grade	Name of Supervisor — Nom du directeur de thèse
1985	Dr. Peter Dixon

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date	Signature
December 7, 1984	Jo Anne LeFevre

NL-91 (4/77)

THE UNIVERSITY OF ALBERTA

A MODEL OF THE USE OF INSTRUCTION AND EXAMPLE INFORMATION ON A SIMPLE INDUCTIVE-REASONING TASK

by

JO-ANNE LeFEVRE

(C)

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE MASTER OF SCIENCE

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1985

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Jo-Anne LeFevre

TITLE OF THESIS: A Model of the Use of Instruction and Example Information on a Simple Inductive-Reasoning Task

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Science

YEAR THIS DEGREE GRANTED: 1985

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed)
Jo-Anne LeFevre

PERMANENT ADDRESS:
305 - 10620 - 79 Avenue
Edmonton, Alberta T6E 1S1

DATED Dec 7 1984

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled A Model of the Use of Instruction and Example Information on a Simple Inductive-Reasoning Task submitted by Jo-Anne LeFevre in partial fulfilment of the requirements for the degree of Master of Science.

Supervisor

Date Dec 7 1984

ABSTRACT

A model is proposed to account for differences in the use of written instruction and specific example information for a simple inductive-reasoning test. Instruction use is defined as the retrieval of a known schematic procedure. This procedure is subsequently instantiated to deal with the particular task being solved. Example use involves the operation of an analogical mechanism that constructs a generalized procedure through a mapping and comparison process. The results of an experiment were generally favourable to the model. Similarity of the test problems to the example was related to accurate performance for subjects who were given an example. During the test, subjects referred to an example sooner and more frequently than to instructions. Furthermore, inductive-reasoning ability was related to successful performance for subjects who were given examples but not for subjects who were given instructions. These results supported the analogical and generalization mechanisms proposed for example use. In contrast, instruction use seemed to depend on how easily subjects were able to retrieve a procedure with the given information. Furthermore, there was some indication that reading speed was related to instruction but not example use, which suggested the processes involved were different. Finally, implications for using this type of instruction and example information are discussed, and the model is extended and refined.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee: Jeff Bisanz, Renee Elio, Alinda Friedman, and Steve Hunka. Their encouragement and erudite suggestions were appreciated throughout the process that culminated in this thesis.

Pete Dixon deserves special thanks, for as they say, "without him, none of this would have been possible". I have learned a lot about designing experiments and analyzing data with his guidance, and together, we have learned what is involved in acquiring a Master's degree.

I would also like to thank Linda Mrkonjic for her help with scoring the protocol data, and for her patience when I was running subjects. I am also very grateful to Margolee Horn for drawing the figures. The anonymous 72 people who participated in this study also deserve some credit. Many of them showed an interest in the hypotheses and results that was very encouraging to me.

Finally, I would like to thank Cathy Dick and Chris Herdman (my fellow survivors) for their patience and support.

TABLE OF CONTENTS

CHAPTER PAGE

I. Introduction 1

Background Research 2

Research on Written Instructions 2

Research on Instructions and Examples 3

A Model of Instruction and Example Use 9

Instruction Procedures 10

Example Procedures 12

Hypotheses 19

II. Method 23

Materials 23

Procedure 27

Subjects 30

Analysis 30

III. Results 33

Part 1: Patterns of Information Use 33

Part 2: Consistent Procedure Use 38

Part 3: Detailed Performance Measures 43

Part 4: Other Dependent Measures 56

IV. Discussion 73

Bibliography 81

Appendix 1. Series-Classification Test 84

Appendix 2. Supplementary One-Way Analyses 88

LIST OF TABLES

Table	Description	Page
1.	The Proportion of Consistent Subjects Using Each Information Source in 6 Conditions from LeFevre and Dixon (1984)	7
2.	The Inductive Problems Used on the Computer Test	24
3.	Metacognitive Questionnaire: Instruction Version	28
4.	Introductory Information for the Computer Test	29
5.	The Instructions and Examples Used in Each Condition	31
6.	Correlations Between Median Reading Speed and the Time Spent on the Instructions or Examples	36
7.	Percentages of Subjects Using Each Type of Procedure as a Function of Information	39
8.	The Percentage of Example Subjects Who Used a Consistent Series Procedure	40
9.	Mean Solution Time and Mean Accuracy Values for the Four Problem Types	45
10.	Mean Solution Time and Mean Accuracy Values for Similarity Analyses	47
11.	Mean Solution Time as a Function of Practice and Information Type	49
12.	Mean Accuracy as a Function of Practice and Information Type	55
13.	Comparison of Consistent Series Procedure Use to Judged Procedure Use	57
14.	Frequency of Specific Examples Given in Procedure Protocols	59
15.	Examples of Procedure Protocols	61
16.	Length, Generality, and Goodness Means Across Problem Types	63
17.	The Percentage of Consistent Computer Subjects Who Used a Consistent Series Procedure on the Series-Classification Test	67
18.	Mean Values on the Metacognitive Questionnaire as a Function of Information and Procedure Use	69

LIST OF FIGURES

Figure	Description	Page
1.	Patterns of Information Use	34
2.	Practice Effects for Schematic Instructions and Specific Example Conditions	51
3.	Practice Effects for Accuracy by Problem Type	52
4.	Practice Effects for Accuracy as a Function of Information	54
5.	Goodness Rating as a Function of Information and Procedure Use	64
6.	Number Series Test Score as a Function of Information and Procedure Use	72

I. Introduction

Many tasks and situations involve presenting some form of instructional information to the person attempting the task. The act of following instructions is important in domains as diverse as filling in forms (Barnard, Wright & Wilcox, 1979), effectively using consumer products (Wright, 1981) and performing well in school (Resnick & Glaser, 1976). It has been suggested that a mental representation in the form of a procedure is constructed from the instructional information and that this procedure is then used to perform the task (Anderson, Farrell, & Sauers, 1984; Anderson, Greeno, Kline & Neves, 1981; Dixon, in press). There are at least two important reasons for studying procedure construction and representation. First, from a practical standpoint, it is important to know the optimal way of presenting instructional information so that a consistent and useful procedure can be created. Second, research on instructions can illuminate aspects of mental representation. Consequently, this thesis focused on two issues: the way in which different forms of instructional information are interpreted and used, and the structure and representation of the resultant procedures.

It is likely that performance on a task that involves instructions is a function of two distinct steps (Simon & Hayes, 1976). First, a procedure for doing the task must be created with reference to the instructional information and prior knowledge. Second, this constructed procedure must be applied to the actual task (Dixon, in press). The level of performance that is achieved through this process is a result of the success of each of the component steps. That is, if the person doing the task fails to extract an adequate procedure from the instruction information, it is unlikely that he or she will be able to perform the task. However, if the procedure is adequate, but applying it to the task taxes some other resource such as memory, success on the task will not necessarily reflect the adequacy of the procedure that has been constructed (Simon & Hayes, 1976). Usually, research concentrates on subjects' ability to apply the procedure they have constructed to the task. Performance is therefore assumed to reflect the execution of the procedure (Sternberg & Gardner, 1983). In this thesis, the application or execution component was made very straightforward. Differences in performance were

therefore assumed to be a function of differences in the procedure construction component. Procedure construction was manipulated by varying the form and presentation of the instructional information.

Two types of instructional information were examined. The first type were "written instructions". For the purposes of this thesis, written instructions were defined as a series of sentences that constituted a generalized or abstract description of how a task should be done. Thus, written instructions either simply named a relevant procedure, or they outlined the processes or steps involved in a possible procedure. The second form of information that was studied was an example. Examples are specific instances that represent the products of the successful application of a procedure. These two components, instructions and examples, together will be refered to as an "instruction set".

The type of instruction sets described above are often used with I.Q. and aptitude tests, on tax forms, in psychology experiments, in computer manuals, and with games. The written component of an instruction set may include an explanation of the task, a procedure for doing the task, and/or a set of rules. An example can vary in how representative it is of the task, and how explicit, detailed, or specific it is. In sum, instruction sets vary in the amount of information they provide, how closely the given information applies to the task, the specificity of the information, and the degree to which a procedure is described.

Background Research

Research on Written Instructions. Much of the research on the comprehension of instructions has focused on aspects of text-processing or reading, or approached the issue from an applied perspective (reviewed in Wright, 1978). One of the most consistent results of this research has been that the form and presentation of instruction information has important effects on performance. For instance, Smith and Goodman (1982) found that hierarchical instructions (those that included reasons for actions and goal information) were read faster and executed with fewer errors than the same instructions presented without this extra information. Dixon (1982)

found that simple action-first directions were read faster than condition-first directions. Jones (1966) demonstrated that phrasing an instruction in a negative way causes subjects to do the task more slowly and make more errors than if the instructions were phrased positively, although the task was essentially the same under both conditions. Results of studies like these suggest that the form of written instruction information is important in at least two ways: it can affect the speed and accuracy of performance, and it can affect the construction and structure of the procedure that is used to do the task.

Recently, Dixon (in press) has proposed a model of the comprehension of written instruction information. In his model, an instruction procedure is represented as a hierarchical structure of actions that are necessary to perform a particular task. When subjects read instructions, stored procedures are activated that can be assembled into a plan for doing the task. These procedures are schematic: that is, they are general versions that contain variable slots for specific information. The slots are instantiated to deal with the particular task being performed. Although this model is useful for understanding how people might deal with written instructions, it does not address the issue of how people use example information, or what the differences between the use of instruction and example information might be.

Research on Instructions and Examples. One of the few extensively investigated tasks for which instruction sets are used is the I.Q. or aptitude test. Most of the aptitude test research concerns the application of a previously constructed procedure to a familiar or "entrenched" task (Sternberg, 1981). Therefore, performance on I.Q. tests is assumed to vary primarily with the difficulty of procedure execution (Sternberg & Gardner, 1983). The difficulty of executing the procedure usually depends on the complexity of the problems (e.g., the number of features that vary along a dimension might increase the complexity of geometric analogy problems and make the analogy more difficult to solve correctly). More recently however, researchers have suggested that the process of constructing a procedure from the task instructions may also be an important aspect of I.Q. test performance (Sternberg, 1981; Sternberg & Ketron, 1983). For example, Resnick and

Glaser (1976) argued "that intelligence can be viewed as the ability to acquire new behavior in the absence of direct or complete instruction" (p. 227). There is evidence for a relationship between the ability to formulate a useful procedure from incomplete instructions and performance on an I.Q. test (Bisanz, Bisanz, & LeFevre, 1984). Some children were unable to use incomplete instructional information that was provided to construct a task-relevant procedure for an analogy task. By collecting protocols and observing the pattern of responses, it was evident that the children had constructed procedures, however, the procedures were inappropriate for the task. These children also had lower scores on another analogy test. Therefore, in these types of tasks, the construction of a procedure may be as important to successful performance as the actual application of the resultant procedure. Although there has not been any research on whether instructions or examples are actually more useful in this context, the instruction sets provided for I.Q. tests usually rely heavily on specific example information and do not include much written explanation.

The role of instructional information has been studied in the area of learning computer programming. Anderson, Farrell and Sauers (1983, 1984) reported that novice LISP programmers rely almost exclusively on specific examples, disregarding instruction information. In fact, Anderson et al. (1983) stated that "Subjects are not able to read instructions, of even modest complexity, and then generate, without error, the behavior instructed" (p. 5). Anderson et al. (1983) suggested that this observation could be traced to the fact that instructions must be converted into procedures before they can be used, and the conversion process is indirect and potentially difficult. According to their model, subjects used worked-out examples as templates for creating new LISP programs. The specific examples provided models from which procedures were more easily extracted than from text, or from which parts of procedures could be taken. The results of Anderson et al.'s (1983) research suggest that example information will be more useful than instruction information, at least on relatively complex tasks like writing computer programs.

A recent series of studies by Gick and Holyoak (1983) also emphasized the importance of examples. Subjects were presented with an example of a successful

procedure in the context of a story evaluation task. First they read two stories in which a protagonist successfully accomplished a goal by using a particular procedure. The stories were either from very similar or dissimilar domains, although the underlying solution, or goal, was similar. Next, subjects were asked to write a short paragraph detailing the similarity between the stories. Gick and Holyoak used this abstract similarity statement to infer the quality of the schema that was abstracted from the two stories. "Good" schemata included the two major points of overlap between the stories and indicated that subjects had understood and noted the underlying analogy. Finally, the subjects were given a problem to solve. The procedure that had been used in the story was useful for solving the problem if the subjects were able to make the connection from the story to the problem-solving task, that is, if the story could be used as an example of a successful problem solution.

Gick and Holyoak (1983) report two important results. First, the greatest amount of transfer from the story to the problem-solving task was found when two prior examples (stories) from dissimilar domains were given to the subjects in the first part of the experiment. Second, the quality of the schema or procedure that was abstracted from these two examples was an important predictor of performance in the problem solving situation. Gick and Holyoak (1983) presented a framework for the use of example information in which an analogical comparison of the example to the new situation is made. This comparison process results in the construction of an abstracted schema that captures the relevant similarity between the stories, disregarding the unimportant points of difference (e.g., setting information). The schema can be applied to the problem-solving task to facilitate problem solution, since the required solution is analogous to the gist of the story solution. The use of examples to construct procedures may operate in a similar way to produce a more general version of a specific procedure.

Rumelhart and Norman (1981) describe a study that helps to define the extent to which example information may be useful. They taught students to use a text-editing system on a computer. Some of the commands could be successfully

taught by providing specific examples. The students were able to construct procedures from some of the examples. However, real progress in understanding the system and using it effectively was not achieved until the students were provided with simple models of how the computer operated. These models involved knowledge the students already had about secretaries, card-filing systems and so on. Specific examples may have been useful for producing immediate performance in this situation, but real understanding of the task and abstraction of task-relevant procedures were more likely to occur when the task was related to prior knowledge and well-practiced procedures.

In previous research the relative usefulness of instruction and example information was examined (LeFevre & Dixon, 1984). The most important influence on procedure construction in the reported sequence of studies was a specific example. LeFevre and Dixon (1984) used a figural inductive-reasoning task that combined the instruction and example information for two different procedures. On each problem, subjects were presented with a series of three figures (e.g., three triangles that decreased in size) and given a choice of answer alternatives. If the subjects used a series completion procedure, they should have chosen the answer alternative that completed the series (e.g., a small circle would complete the decreasing size series). The other procedure was a similarity procedure, in which the subject should have chosen the answer alternative that was most similar in appearance to the three figures (e.g., a triangle). Answers for the two procedures were mutually exclusive. The goal of this study was to examine the relative usefulness of instructions and an example when these two sources of information described conflicting procedures.

When subjects were given general instructions that described one procedure and a specific example that demonstrated the other procedure, most of them used the example procedure to do the task. The instructions were essentially ignored. This "example effect" was consistent across a wide variety of manipulations. Table 1 presents the percentage of subjects who used each procedure (the instruction procedure or the example procedure) consistently for each experimental manipulation.

Table 1

The Proportion of Consistent Subjects Using Each Information Source in 6 Conditions from LeFevre and Dixon (1984)

Condition	Followed Instructions	Followed Example	p value (sign test)
1. Size Example	.08	.92	.001
2. Number Change Example	.07	.93	.001
3. Example first	.07	.93	.001
4. Counter-information	.25	.75	.038
5. Long Instructions	.32	.68	.054
6. Instruction Emphasis	.31	.69	.105

The most uninteresting and obvious explanations for the example effect were quickly ruled out. The example effect was not specific to a particular example. A size-change example (previously described) and an example in which the number of figures decreased or increased were used equally often (conditions 1 and 2 in Table 1). The example effect was also not due to the order in which instruction and example information was presented to the subjects (compare conditions 1 and 3 in Table 1). The effect was just as strong whether instruction or example information appeared last. A number of manipulations were designed to improve or emphasize the written instructions. Information about incorrect alternatives was added (counter-information, condition 4), the instruction information was made more redundant and procedural (condition 5), and the experimenter verbally stressed the importance of the instructions (condition 6). However, the data in Table 1 indicate that these manipulations attenuated but did not eliminate the example effect. The conclusion from these results was that examples are an important source of information on this particular task.

One obvious explanation of the example effect is that the instructions that were provided for this task were simply not very good. If this were the case, then people might be forced to rely on the example when they were given conflicting information, since they could not interpret the instructions. As Anderson et al. (1983) suggested, subjects might find written instructions very difficult or impossible to use when constructing a procedure. However, in a condition in which subjects were given only one source of information (instructions or an example), approximately the same percentage were able to use the information provided to construct a consistent procedure (54 and 52 percent, respectively). So, although neither source of information alone seems to be extremely useful for this task, both sources <u>can</u> be used successfully by about half the subjects.

Another explanation of the example effect offered by LeFevre and Dixon (1984) was that subjects ignore instructions because they do not believe that the instructions will be very useful. This explanation is consistent with research reported by Barnard et al. (1979) and Wright (1981). Barnard et al. found that subjects

tend to ignore the instructions for a simple task (filling in a form), possibly because the subjects believed that they already knew what to do. Similarily, Wright (1981) found that people ignore the instructions given for common appliances because they believe that they know how to operate these familiar appliances and therefore regard the instructions as unnecessary. Protocols collected from subjects in one experiment of LeFevre and Dixon (1984) were consistent with this explanation. When subjects were asked why they used the example, some subjects mentioned that instructions are not very useful, and most subjects said they thought that examples are usually clearer and easier to understand than instructions. LeFevre and Dixon (1984) concluded that the example effect may be an effect of prior experience with instructions and examples: subjects believe that instructions are not useful, that they are difficult to interpret, and/or that they are unimportant. However, there was no compelling evidence that the example effect could not also represent some kind of processing advantage. For instance, examples may provide templates from which procedures are quickly and easily extracted (Anderson et. al,. 1983), reducing processing time or effort.

A Model of Instruction and Example Use

The research described in this thesis was designed to (a) examine the possibility that the structure, presentation, and content of instruction and example information is an important influence on the construction of procedures and (b) investigate how the procedures constructed from the two sources of information might differ. A two part model of the use of instruction and example information was formulated in order to highlight differences in the use of instruction and example information. This model is intended as a framework for guiding research, and so the study done in this thesis will not necessarily pertain to all aspects of the model. However, it is hoped that the framework will prove useful for organizing a wide range of data. In this section, situations in which different procedural representations are created from example and instruction information are described, and processes are proposed that may be involved in the development of these procedures.

The proposed model of instruction and example use required a few assumptions. First, the model only applies to the situation in which one source of information is available, either instructions or an example. Second, the instructions described by the model are short and simple ones that outline a task rather than explain it in detail. Third, only one example is available to the subject. In this restricted situation, two different mechanisms are proposed: one to deal with written instructions, and the other to deal with specific examples. The procedures that result from the operation of these mechanisms are distinct in structure and content. Therefore, performance differs both as a function of the differences in the way procedures are constructed, and as a function of differences in the resulting representations.

Instruction Procedures. In the proposed model, the presentation of simple written instruction information results in the retrieval of a stored procedure. The words in the instructions are used to retrieve previously constructed and stored schematic procedures that can be specified or adapted to deal with the task at hand. Recipes provide good instances of this sort of instructions. For example, "Saute the vegetables" is a short and simple instruction, but the cook retrieves a whole series of actions with the word "saute" and is able to successfully perform the instruction. As described previously, Dixon (in press) provides a coherent description of the hierarchical nature of instruction information. According to his model, plans that are created from the task instructions are hierarchical collections of related actions. The most general or all-inclusive actions form the higher levels of the hierarchy, and the actions become more specific and detailed at lower levels. Therefore, the level of specification of the instruction becomes an important variable, and the generality of the retrieved procedure will vary with both the task and the person doing the task. In the above example, specifying "saute the vegetables" as "cook the vegetables over medium heat in a small amount of oil until soft, stirring frequently" would be expanding the instruction to include actions from a lower level of the hierarchy. Thus, one source of difficulty with instructions may be their level of specification. The most appropriate level varies with the purpose and the population to which the

instruction applies. Anderson et al.'s (1983) concern with the difficulty of interpreting instructions for programming in LISP probably only applies to novice programmers. It is likely that more experienced LISP programmers might find the same instructions quite useful and comprehensible, since they are able to access the component actions and relevant goals.

In sum, the proposed model of instruction use includes four major assumptions. First, the model does not deal with the situation in which a new procedure is created although a possible method would be to use old procedures to construct new ones. Second, simple written instruction information is not used to construct a procedure, but instead is used in conjunction with previous knowledge to reference and retrieve a procedure. Third, this retrieved procedure is instantiated to deal with the current task situation. Fourth, the procedure will not undergo much change in the course of the task if it was adequate for dealing with the situation. However, if an applicable procedure is not retrieved with the help of the information in the instructions, then performance may be quite erratic because the person must use other sources of information to try and construct a workable procedure. In general though, for the type of "entrenched" task being studied (Sternberg, 1981), the short instructions that are provided are primarily used to retrieve a solution method. Application or execution of the procedure becomes the major source of difficulty and variability in performance.

This characterization of how instructions are used may also help to explain why people tend to ignore instructions. If the instruction user believes that the instructions are primarily useful for retrieving a known procedure, and, if he or she can retrieve a procedure without reading the instructions, then the user may think the instructions are unnecessary and disregard them. This explanation might help to account for the results obtained by Barnard et al. (1979), Wright (1981) and LeFevre and Dixon (1984). A second implication of this model is that, if subjects read the instructions and then were unable to easily retrieve a procedure, they might then pay more attention to a specific example and disregard the instructions (LeFevre & Dixon, 1984). In other words, meta-knowledge about instructions (and

examples) and their use may also have important effects on performance.

Example Procedures. Contrary to the situation for research on instructions, there are no existing models of example use that could be adapted and expanded to deal with the present case. However, two areas may be helpful for understanding the example situation, and the proposed model will incorporate elements of both. The first is research on analogical processes (Gentner, 1983; Sternberg, 1977) and the second is learning ill-defined categories from examples (Elio & Anderson, 1981; Hayes-Roth, 1978).

A model of example use that incorporates analogical mechanisms seems quite tractable (Anderson et. al, 1981; Gick & Holyoak, 1983; Rumelhart & Norman, 1981). Analogical models have been proposed for a number of areas, including understanding metaphors (Tourangeau & Sternberg, 1981), and as the mechanism underlying modelling as a useful teaching tool (Holyoak, 1984). The common and familiar nature of analogy might help to explain the widely held belief that examples are somehow "better": people can always at least attempt to reason analogically if an example is provided, whereas the use of instructions may be more dependent on prior knowledge.

Gentner (1983) described analogy as "an assertion that a relational structure that normally applies in one domain can be applied in another domain" (p. 156). In her model, the attribute or specific feature information is dropped from the analogy and only the higher level or relational similarities between the items being compared are retained. This process, which Gentner terms "structure-mapping", resembles extracting a schema that summarizes the similarity between two items or domains. Since examples are usually from similar domains as the task to which they apply, and since the relations within those domains should also be relatively similar, this postulated structure-mapping process should operate easily on examples. One obvious implication of the Gentner model is that the less similar the domain and relational structure of the example is that of the task, the harder the mapping process will be. For instance, an example of a solved calculus problem would probably not be very useful for figuring out a problem in matrix algebra. Although

both problems fall within a broad domain of "mathematical problem solving", the structure of the domains and the problem solving processes used are not very similar.

In the area of learning ill-defined categories from examples there are two competing views of what is learned. One view (Brooks, 1978; Medin, Altom, & Murphy, 1984; Medin & Schaffer, 1978) is that specific example information is retained from the learning process and retrieved in order to make category judgements. The other view (Elio & Anderson, 1981, 1984; Hayes-Roth, 1978; Hayes-Roth & Hayes-Roth, 1977) suggests that generalized representations are constructed and used to classify instances, although some specific information may be retained. This controversy is important to the study of examples because the form and content of the representation (specific versus general) has different implications for performance. If only specific information is retained, then an example procedure will never become as general as an instruction procedure and the process needed to use an example is a straightforward analogical comparison. Subjects might simply compare the example, or a memorized representation of that example, to each subsequent problem in the course of solution. On the other hand, in order for a more generalized representation to be produced, some mechanism beyond a simple analogical comparison must be included in the model. This second possibility implies that the resultant procedure will be more like an instruction procedure in terms of generality, although it does not answer the question of how general an example procedure can become. It is possible that example procedures will eventually become as general as instruction procedures, but only after experience with a variety of different tasks for which the same procedure is useful.

Brooks (1978) reported data that supported a model in which specific stored information is used to determine category membership. According to his model, general category descriptions (e.g., schemata) were not extracted from a series of specific examples. In another concept learning study, Medin, Altom, and Murphy (1984) found that subjects who were given only category exemplars from which to learn a concept based their judgments of category membership on stored exemplar

information. However, subjects presented with either prototype information or a combination of prototype information and specific exemplars used both prototype and exemplar information to make judgments. These studies suggest that subjects given only examples will not produce a representation that is generalized. Instead, they will rely on specific information to perform the task.

In contrast, Elio and Anderson (1981) proposed a theory of concept learning or schema abstraction in which generalized structures were created in response to example information. In their generalization model, category descriptions are formed as a subject experiences a variety of instances from the category. The constant or defining features across instances are retained and strengthened in the generalization each time they reoccur. The features that vary across instances are represented by slots in the generalization that can take on a variety of values. This characterization may be very similar to the way in which examples are used to construct a procedure. A comparison is made in order to extract the overlapping or common features (where features can be parts of procedures as well as attribute information) and these commonalities form the basis of the procedure. However, in the experiments that Elio and Anderson (1981) reported, they also found evidence for a similarity effect, suggesting that some sort of specific instance information is retained and used. In the case of examples, this specific information may be retained in the form of possible instantiations of variable slots. If a problem is very similar to a preceding example which formed a partial basis for a generalization, it is likely to easily be solvable by the generalized procedure.

In the model of example use that is proposed, both analogical comparison and generalization are postulated to occur. Comparison processes initially operate on specific instances (the example and the current problem). However, as the generalization is developed, the problem is compared to the generalization and not to the specific example. Specific information is retained in the form of candidate instantiations of slots in the generalized procedure. A concrete example of procedure construction may clarify this process. The example given to subjects in LeFevre and Dixon (1984) consisted of three triangles that decreased in size and three answers, a

triangle, a small circle, and a square. Suppose the correct answer indicated for this example was the small circle. If the first problem consisted of three squares that increased in size, the generalization created by comparing the problem to the example might be: "a problem is three figures (triangles or squares), the size of the figures changes (increase or decrease), and the answer is the same type of size change (a larger or smaller figure)". The items in brackets represent slots whose values can vary, while the other statements capture the commonalities between the problems. If the third problem had three circles decreasing in size, this generalized procedure should easily find the correct answer. If the third problem was a series of groups of circles, first one circle, then two circles, then three circles, a comparison of the problem to the generalization should result in an augmented generalization, such that: "a problem is three groups of figures (where group can be one or more), there is some kind of change (size (increase, decrease) or number (increase)), and the answer figure also changes correspondingly (size (smaller, larger) or number (more)). The final procedure produced by this process would therefore be generalized and yet task- or problem-specific to some degree. Difficult problems would be those in which there was little similarity evident between the problem and the generalization. In the task described for the above example, a difficult problem was one in which the number of sides of each of the figures changed. Evidently, this change was not very similar to the other changes that occured and subjects had difficulty using their procedures to select the appropriate answer.

In sum, procedure construction in the proposed model proceeds as follows. First, a representation of the example that will serve as the basis for the constructed procedure is created. Second, the example or its representation is compared to the first problem and the similarity between the two is incorporated into a generalization. Subsequently, each problem is compared to the generalization (or in rare cases back to the example if the similarity between generalization and problem is difficult to extract), and the generalization is modified to deal with each problem. Although attribute similarity may help the comparision process by reducing the number of irrelevant dimensions that conflict, the important similarities will be

those that are relevant to generalization. In the above example, the important and notable similarity across problems is that something (be it size or number) is always changing.

The procedure that is constructed with this generalization process will be continually refined, and yet it will only be as flexible as the problems allow. Since there is often no feedback given after the initial example, inaccuracies may get permanently incorporated into the procedure. Furthermore, the procedure probably does not contain much information from abstracted levels of the procedure hierarchy that would be standard for an instruction procedure. In this sense, instruction-generated procedures may be more "general" than example procedures because instruction procedures may be potentially applicable to a variety of different tasks. For instance, a person who used examples of statistical analyses to construct his or her own analyses might always get correct results and become quite proficient. However, it is likely that the person would not understand why certain steps were necessary in the analysis, and would only continue to include them because it worked before. If an analysis was significantly different from previous ones, it is likely that the lack of more general information would become evident. The person may not know which particular steps were relevant to any part of the procedure and so be unsure about which steps to modify for the new analysis. Unless an example-generated procedure could be expanded to include information from levels of the hierarchy that have organizing and extending information, the usefulness of the procedure may be limited. Rumelhart and Norman (1981) noted this deficiency for their subjects who were learning to use a text-editing system, and it is probably also true for subjects learning LISP if those subjects did not also read the explanatory text provided with the examples (Anderson et al., 1983). Anderson et al. (1981) also noted that examples of geometry proofs were often not sufficient to help subjects construct subsequent proofs, and could even be misleading when the similarity between the example and the new proof was only superficial and not substantive. Since performance is generally better when subjects are given high level, explanatory information (Dixon, in press; Smith & Goodman, 1982), this

characteristic of example procedures is a potentially limiting factor.

It is not clear from the preceding description how general or schematic an example-generated procedure can become. An example procedure may only be generalized to the extent that it is not problem-specific, although it may well continue to be task-specific and not easily articulated or explained in the same way that an instruction-generated procedure might be. For instance, subjects may not be completely aware of the actual dimensions that are important in the task, although they may be able to respond correctly on the basis of their knowledge of the specific instances encountered (Brooks, 1978). In this sense, the generalization process may be largely unconscious and the procedure that is created therefore may not be easily verbalized. Another way of phrasing this is that subjects may not know the reasons that they perform any part of their procedure, although they may invariably produce the correct response. Depending on what subjects are supposed to gain from doing the task, this may or may not be detrimental to their future performance (cf. Rumelhart & Norman, 1981).

A prediction of this generalization model that is alluded to in some of the papers discussed (Elio & Anderson, 1981; Gentner, 1983) is that the similarity of the example to the task will be very important in a number of ways. The definition or conception of similarity that will be used follows that described by Tversky (1977) of overlapping feature sets, the important features being defined by the context of the intended application and by the direction of the comparison. However, "feature" is used in a broad sense to include attributes, relational features (Gentner, 1983), and between- and within-domain correspondence (Tourangeau & Sternberg, 1981). Practically, the perceived similarity of an example to a task or item is probably quite dependent on the context of the task. For instance, a recipe for chocolate cake will be very similar to one for white cake if the domain is that of recipes, but if the domain was that of dessert flavours, chocolate and white cakes would be considered considerably less similar. Therefore, the similarity of any given example to its task may depend on one or two important but context-dependent dimensions. In the LeFevre and Dixon (1984) study discussed

earlier, similarity was defined after the fact as the type of change involved in the problem (i.e., size, number or orientation), because this dimension seemed most salient within the task.

Similarity has important implications for how useful an example will be. Problems or tasks that are very similar to the example (whether in terms of the rule involved, the specific attributes of the problem, the domain, or the relational structure) will be very easy to perform. The elements of the problem will easily map on to the example and to the generalization, since the generalization is at least partially based on the original example. Therefore, performance should be more accurate for problems that are similar to the example than for problems that are not similar to the example. However, similarity may also retard the generalization process. Exposure to problems in which the domain and the specific attributes are different from the example seems necessary for the relational similarities to be highlighted and included in the procedure (Elio & Anderson, 1984; Gentner, 1983; Gick & Holyoak, 1983). Furthermore, if the relational structures are too dissimilar, no mapping of relational structure may be possible, and the example may be quite useless (e.g., calculus and matrix algebra). The relative similarity of the example to the task as a whole may determine how "good" an example it is. Consonant with this prediction, in the data reported by LeFevre and Dixon (1984), the similarity of the example to the problems was a consistent determinant of performance even for subjects who could not construct a procedure that was relevant to the whole test. These subjects were significantly more likely to do the problems that resembled the example with the procedure that the example modelled. However, they did not seem to extract a procedure that could be applied to a wide variety of problem types.

The above descriptions of example and instruction information suggest that the procedural representations that are formed and the process of forming them would be quite different with the two different sources of information. In comparison to an example-generated procedure, an instruction-generated procedure would include information from a higher hierarchical level and would not be as closely tied to specific information as an example-generated procedure. Performance should be fairly

stable across the test as a whole, assuming successful retrieval, and the final procedure arrived at by the subject would be relatively well articulated (subjects could label and describe it). In contrast, the procedure generated by the example information would contain more specific information about the type of problems on the test, and the actual procedure would not be well specified in terms of a verbal description. Especially at first, an example procedure would be affected both by variations from problem to problem and by the similarity of the problems to the example.

In sum, the proposed model of instruction and example use suggests that instructions provide a retrievable schema (an abstract description) that can be expanded and detailed into a task-applicable procedure. Knowledge about the kind of task being attempted is used in conjunction with a set of applicable procedures to produce a procedure that is task-relevant and yet remains fairly general. Conversely, examples illustrate the result of a successful procedure. They are used with both comparison and generalization mechanisms, the final result being a constructed procedure that has been generalized in the process of comparing test problems to the constructed representation and generating the best answers on the basis of that comparison. The main difference between the instruction and example situation is that instructions provide the outline of a procedure that the subject must make specific enough to apply to the task, whereas the example is already very specific, and the subject must generalize. This difference introduces at least one possible source of subject variability, since some subjects might find the inductive use of the example most compatible with their abilities, and other subjects might be better at the deductive nature of instructions, or perhaps have better memories or more knowledge and find it easier to retrieve appropriate procedures.

Hypotheses

In this thesis, several hypotheses were tested that derive from the model just described, and from the assumptions that were made about the type of information that instructions and examples provide. These hypotheses will be explained within the context of the experimental procedure that was used. In brief, problems were

presented to subjects on the CRT screen of a computer. Dependent measures included: the time spent on the problems, the answer chosen for each problem, and the time spent reading instruction or example information. Subjects were also given the opportunity to look back at the instruction or example information, and the pattern and time-course of this "looking back" was recorded. Subjects in this study received only one source of information, either written instructions or an example. Therefore they were forced to use this information source as best they could in order to construct a procedure for solving the problems.

Two forms of both the instructions and examples were used. Instructions were presented in either a "procedural" or a "schematic" form. Schematic instructions provided the label of a stored procedure, whereas procedural instructions included the steps of a possible procedure at a lower level, without specifying the higher-level label. (The "washing clothes" paragraph used by Bransford and Johnson (1972) is an example of providing a procedure without the label.) Examples were presented in either a "specific" or a "prose" form. A specific example was an instance of one of the problems that could be encountered on the test with the correct answer indicated. The specific example resembled the test problems physically. It included the incorrect answer alternatives or counter-information. The prose example summarized the information a subject might obtain from a specific example, without including information about incorrect alternatives. It was essentially an outline or description of a particular type of problem one might encounter on the test.

The first hypothesis was a direct implication of the proposed model of example use. If examples are used in a comparative way, problems that are similar to the example should take less time to solve because the comparison process is more straightforward. Similar problems should also be solved more accurately than problems that are not like the example. Very little extra generalization would be necessary for similar problems, since the generalized procedure could simply be applied to the problem. Therefore, the similarity effect should be consistent across the test, although it may be more important at the beginning of the test when

procedure construction has just started.

The second hypothesis is that instruction users will show less of a practice effect because they are applying rather than constructing a procedure. Instruction users retrieve and instantiate a procedure, so that their performance becomes stable quickly. Example users are constructing a procedure and will therefore be more influenced by individual problem variation. For example users, construction will be replaced by application after some amount of experience with the task.

The third hypothesis is that people who use examples will refer back to the example more frequently, and may spend less time on it initially. Instruction users should spend more time initially perusing the instruction information and not refer back as much. This hypothesis stems from the model of the use of instruction and example information. Examples are initially used in a comparative way, so therefore more looking back during the first few problems may be necessary. Instructions are used to retrieve a previously constructed procedure that is instantiated by experience with the problems.

A fourth hypothesis is that the procedures articulated by example and instruction subjects will be different. Example users may be more likely to give a procedure that relies on specific information, whereas instruction users may summarize or name their procedures. The reported procedures may reflect the extent of the procedure hierarchy and composition; example procedures may contain specific steps, and instruction procedures may contain more general actions. The quality of the reported procedure will be reflected in the subject's ability to transfer their procedure to a related task. Example procedures will be less readily generalized than instruction procedures to a task that uses a similar procedure in a different domain since a higher level procedure is not necessarily created, and the example-generated procedure may be quite sensitive to individual problem characteristics.

Overall performance on the experimental task is not expected to differ greatly with the type of information that is provided. However, the pattern of performance across problems will be used to show (a) that instructions and examples are used differently, even in a relatively simple situation, and (b) that the procedures

constructed in response to the different types of information vary in both scope and structure.

II. Method

The study described in this paper was an analysis of how people interpreted instructions and examples, and how this interpretation was reflected in their performance. A task was designed in which the application component of performance (e.g., using the procedure to perform the task) would be fairly simple, so that differences in performance could be attributed to the procedure construction component of instruction or example use. Different forms of instruction and example information were given to subjects and then performance on a number of dependent measures was analyzed. The results of this study should help to answer questions about how examples and instructions are used to construct procedures.

Materials

Computer Task. The primary task that was used in this study was a modified inductive-reasoning task. It was essentially a simple series-completion test. Each question on the test consisted of three words (the "problem words") and three answer alternatives. The problem words came from the same category, and formed a series (e.g., "mouse dog horse" are mammals that increase in size). The correct answer for this problem would be a larger mammal, "elephant" for instance. The change that occured across the problem words should have been readily apparent to the subjects. For example, for the problem words **December, February, April**, the correct series answer would be **June**. Two distractor alternatives were also included. One was the name of the category that the problem words belonged to, and the other was a word from the same category that did not complete the series. This distractor was included to check the precision of the subjects' series procedures. It indicated whether subjects were simply choosing a related category exemplar or if they were actually using a series completion procedure.

There are four classes of problems, as shown in Table 2. These include: animals with size changing (e. g., "rhinocerous cow wolf"), measures with amount changing (e. g., "gallon quart pint"), ordinal series (e. g., "Monday Wednesday Friday"), and other problems that represented a variety of transformations (e. g., "plains hills mountains"). The purpose of having well-defined classes of problems was so that the similarity of the example to individual problems could be easily

Table 2

The Inductive Problems Used on The Computer Test

Class 1: Animals, size changes			
mouse dog horse	A. mammals	B. elephant	C. raccoon
rhinocerous cow wolf	A. rat	B. zebra	C. mammals
St.Bernard collie spaniel	A. husky	B. dogs	C. pekinese
toy-poodle terrier dalmation	A. yorkie	B. German shepherd	C. dogs
lizard rattlesnake alligator	A. dinosaur	B. reptiles	C. salamander
crocodile cobra garter-snake	A. reptiles	B. boa constrictor	C. chameleon
goldfish trout halibut	A. shark	B. minnow	C. fish
swordfish salmon angelfish	A. guppy	B. shark	C. fish
chickadee hawk vulture	A. sparrow	B. birds	C. ostrich
eagle duck robin	A. birds	B. hummingbird	C. flamingo
Class 2: Measures, amount changes			
gallon quart pint	A. volume	B. cup	C. bushel
litre decilitre centilitre	A. millilitre	B. kilolitre	C. volume
milligram centigram decigram	A. microgram	B. weight	C. gram
ton half-ton pound	A. semi-ton	B. ounce	C. weight
inches feet yards	A. miles	B. length	C. half-inches
kilometers meters centimeters	A. length	B. decameters	C. millimeters
second minute hour	A. day	B. millisecond	C. time
decade year month	A. time	B. weeks	C. century
1/8-cup 1/4-cup 3/8-cup	A. 3/16-cup	B. fractions	C. 1/2-cup
2/3-pound 1/2-pound 1/3-pound	A. fractions	B. 1/6-pound	C. 5/6-pound

Class 3: Ordinal Series

December February April	A. months	B. June	C. March
October September August	A. July	B. June	C. months
Tuesday Monday Sunday	A. Friday	B. days	C. Saturday
Monday Wednesday Friday	A. Thursday	B. Sunday	C. days
B D F	A. H	B. letters	C. N
Y X W	A. letters	B. T	C. V
ninety-two eighty-nine eighty-six	A. eighty-three	B. eighty-two	C. numbers
seventeen twenty-one twenty-five	A. numbers	B. twenty-nine	C. twenty-seven
summer fall winter	A. autumn	B. seasons	C. spring
autumn summer spring	A. seasons	B. winter	C. fall

Class 4: Other Problems

car tricycle motorcycle	A. vehicles	B. unicycle	C. bus
secretary manager vice-president	A. president	B. receptionist	C. employees
mountain hill prairie	A. peaks	B. landforms	C. valley
rowboat sailboat yacht	A. raft	B. battleship	C. boats
calm breeze storm	A. hurricane	B. weather	C. sunshine
Brazil Mexico U.S.A.	A. countries	B. Argentina	C. Canada
depressed unhappy happy	A. ecstatic	B. sad	C. moods
gorgeous pretty plain	A. appearance	B. ugly	C. handsome
frozen cold warm	A. icy	B. temperature	C. hot

controlled. The Animal, Measure, and Ordinal problems were designed so that those within any class would be more similar to others within that class than to problems in other classes. The two dimensions of similarity that were considered included the category of the words in the series, and the type of change that occured. The Other class of problems were intended primarily as filler items.

The test had a pool of 39 problems, 10 in each of the Animal, Measures, and Ordinal classes and 9 in the Other class. For each subject who received an example, the example was chosen randomly from one of the first three categories. The other 9 problems in that category were used in each subject's version of the actual test, plus 9 randomly chosen from the other three categories for a total of 36 problems.

The composition of the problems in the test was determined according to a scheme described by Kenny and Smith (1980). Three different latin squares or problem combinations were constructed. First, two examples were randomly chosen from each of the three example categories and each was assigned to a subject in each information condition (the instruction subjects never actually saw their examples). The remaining eight problems in each category were included in the body of the test. Four problems from each category were randomly selected for each half of the test. Pairs of subjects then received tests that were balanced in terms of first half/second half composition and type of example. The four remaining problems were put in between the first and second halves of the test. This process was repeated three times, resulting in 18 subjects per condition, six from each of the latin squares. The actual order of the problems within each half was randomly determined for each subject.

Other dependent measures. After the series-completion test was finished, subjects were asked to write a short description of how they solved the problems on the test, as if they were explaining to someone who knew nothing about it. This was termed the "procedure protocol". Next, they were given a short, paper and pencil version of the Series-Categorization task used by LeFevre and Dixon (1984).

This test is included in Appendix 1. It was termed the "Transfer Task" because the series procedure required for the Series-Categorization task is similar to the series procedure required for the Computer Task. The subjects were instructed to "do this test the same way you did the computer test".

Subjects were then given a five item metacognitive questionnaire to answer (the instruction version is shown in Table 3) that was designed to probe their opinions about the use of instruction or example information. Finally, subjects solved the Number Series Subtest of the Lorge-Thorndike Intelligence Test Battery (Lorge & Thorndike, 1954). According to the test makers, this test measures inductive-reasoning ability. The use of an analogical mechanism with examples may be related to performance on the number series test, since both tasks utilize inductive processes. Subjects were given the standard instruction set and nine minutes to work on the 28 item test.

Procedure

The Computer test was administered as follows. First, subjects were seated in front of a CRT screen. A box with labelled buttons on it was positioned in front of the screen within easy reach of the subjects. Subjects pressed an indicated button to begin the presentation of the test. In sequence, the introductory paragraphs shown in Table 4 were presented. Subjects read each paragraph at their own rate, then pressed an indicated button to continue. Reading time for these paragraphs was recorded as a measure of the subject's overall reading speed. This process was also intended to familiarize subjects with the use of the buttons. Then subjects were shown instruction or example information as they held down another button (marked "X" on the button box). This information only stayed on the screen for as long as the button was held down. When a subject finished reading the instruction or example information and released the "X" button, a menu was displayed. The menu had two choices: "Hold down X to see the instructions" and "Press N to see the next problem". This menu also appeared between each pair of problems. If the subject held down the X button, the instruction or example information was displayed. If they pressed N, the next problem appeared on the

Table 3

Metacognitive Questionnaire: Instruction Version

Please answer the following questions about the test that you did on the computer. Circle a number on the scale below each question. Ask the researcher if you don't understand a question.

1. How useful did you find the instruction information for this test? (the instruction information appeared when you held down the button labelled 'X').

1	2	3	4	5
not very useful		somewhat useful		very useful

2. How much did you use the information in the instructions to help you solve the questions on the test?

1	2	3	4	5
a little		some		a lot

3. How useful do you usually find the instructions that are given for tests like this? (That is, tests in which you are given short instructions, with one or two examples of problems, and then are expected to solve a set of varied problems that may or may not be similar to the example).

1	2	3	4	5
not very useful		somewhat useful		very useful

4. How difficult or easy did you find this test?

1	2	3	4	5
very hard		moderate		very easy

5. Suppose you had been given a specific example of a problem on the test instead of written instructions that described how you should solve the problems on the test. How useful would this type of information have been?

1	2	3	4	5
not very useful		somewhat useful		very useful

Table 4

Preliminary Information for Subjects

Screen 1

In this study, you will be answering questions. The questions will be displayed on this screen. You will answer these questions by pressing one of the buttons that appear on the box in front of you. You will also use one of the buttons to move from one question to another. Press the button that is labelled "N" to continue.

Screen 2

The questions that you will be answering will appear in a multiple choice format. You will choose one of three answers. As you can see, three of the buttons are labelled with the letters "A", "B", and "C". You will use these buttons to indicate your answers to the questions. Now, press the button that is labelled "A" to continue.

Screen 3

The button in front of you that is labelled "N" is used for moving from one question to another. The "N" stands for NEXT question. The button that is labelled "X" will be used when you want to view the information that describes how you should solve the questions on the test. Press the button labelled "N" to continue.

Screen 4

In this test, you will see groups of words. Each group is one problem. The first 3 words are the question, and you are given the choice of 3 answer alternatives. You are to pick only one of these answers and press the appropriate button (A, B or C) on the box in front of you. Press the button labelled "N" to continue.

Screen 5

The button that is labelled "X" should be used when you want to view the information that will help you decide how to do the test. AT ANY TIME, if you want to see that information, hold down the button labelled "X". The information will be displayed for as long as you hold down the button. Hold down "X" to display the test information.

screen. At the bottom of the problem screen, the message "Hold down X to see the instructions" appeared. When the subject answered the problem by pressing A, B, or C, the problem disappeared and the menu reappeared.

After subjects finished the Computer test, the other four dependent measures were administered in the order: procedure protocol, transfer test, questionnaire, and number series test. Total time required to complete this study varied from 30 to 50 minutes. Subjects were debriefed and asked for their comments afterwards.

Subjects

In total 72 subjects participated in this study, 18 in each of the 4 conditions. Forty-six of the subjects were introductory psychology students who received course credit for their participation. Twenty-six of the subjects were paid $3.00 each for participating. There were 34 males and 38 females, and subjects were randomly assigned to conditions.

Analysis

The results were analyzed using a hierarchical design. The primary factor was Information, and the two levels of Information were Instructions and Examples. Within each of the Information types, there were two different Forms of information. These four types of Information are shown in Table 5. The Form effect has two degrees of freedom. One is assigned to the effect of Schematic vs Procedural within Instructions, the other to the effect of Prose vs Specific within Examples. Since the comparison between instructions and examples was usually of the most interest, this nested design seemed more appropriate than a one-way analysis with four factors (Winer, 1971). However, the one-way analysis was also done for the major effects. In all cases, the interpretation of the results using a one-way analysis and appropriate contrasts (instructions vs examples) was not different than with the hierarchical design. Therefore, the hierarchical design results are discussed in the Results sections, and the source tables for the one-way analyses are included in Appendix 2.

To assess practice effects, the Trial effect contrasted performance on the first 16 problems to performance on the last 16 problems.

Table 5

The Instructions and Examples Used In Each Condition

Schematic Instructions

The first three words of each problem form a series. Each problem that you will solve forms a particular type of series. There are three answer choices for each problem. You should choose the correct answer. Answer each problem by pressing A, B, or C. The correct answer to a problem is the word that continues the series formed by the first three words.

Procedural Instructions

Read the first three words of each problem. Decide what the change is across the three words. The change will be consistent from one word to the next. The change can be described by a simple rule. Use the rule to decide which of the three answer choices follows from the first three words. Answer each problem by pressing A, B, or C.

Prose Example

Here is an example:

If the first 3 words were mammals, and they were in increasing order of size, then you should choose the large mammal as the correct answer.

Specific Example

Here is an example:

mouse dog horse

A. mammals

B. elephant

C. raccoon

The correct answer is B.

For comparisons that involved the differences between two proportions, a joint-binomial test was used. The following explanation of this test was taken from LeFevre and Dixon (1984). For the joint-binomial test, the difference between two proportions is tested by calculating the sampling distribution of the difference in proportions, $X_1/n_1 - X_2/n_2$. Here X_1 and X_2 are the number of subjects who use a given strategy in conditions 1 and 2, and n_1 and n_2 are the sample sizes. The null hypothesis of no difference in proportions was rejected if the probability of a difference as large as that obtained was smaller than 0.05. Under the null hypothesis X_1 and X_2 have a joint binomial distribution with a common probability parameter: $Pr(X_1,X_2) = B(n_1,X_1,p)\ B(n_2,X_2,p)$, where B is the binomial probability mass function. The parameter p was estimated by $p = (X_1+X_2) / (n_1+n_2)$.

III. Results

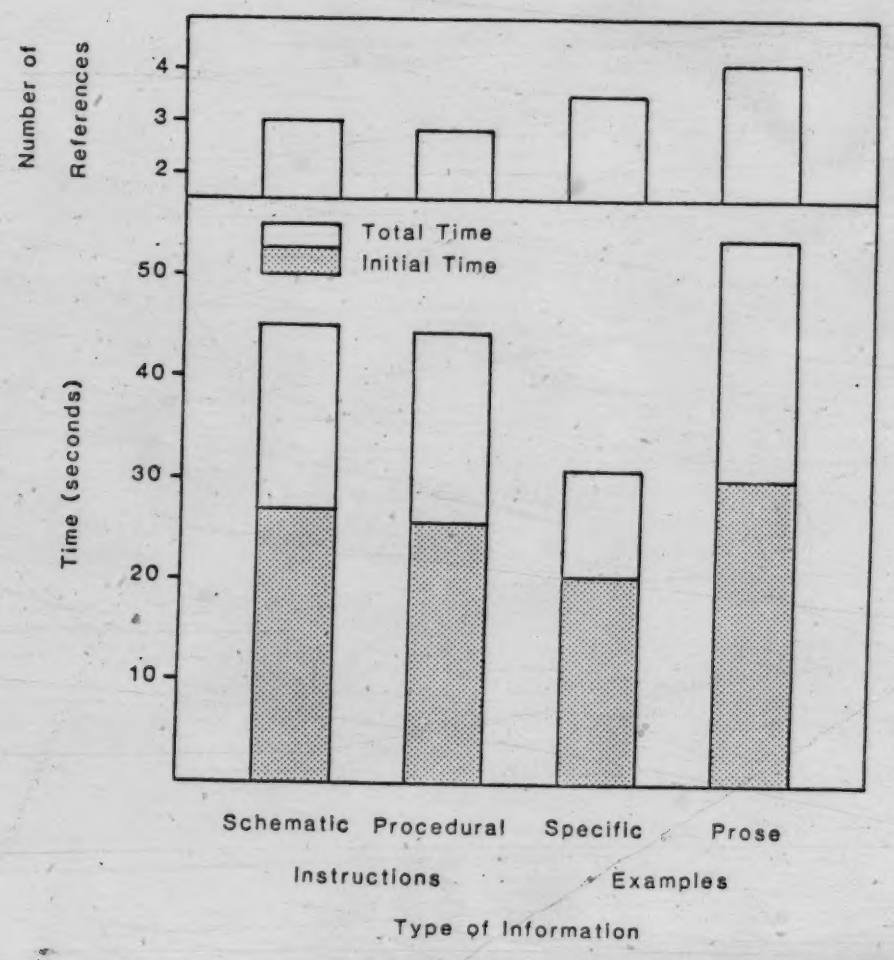
Description of the results is divided into four major sections. In the first section, the pattern of information use (instruction or example) is discussed. This includes analysis of the amount of time spent reading the instructions or example, the total number of references to the instruction or example information, and the pattern of references to the instructions or example. In the second section, consistent procedure use on the Computer test (overall performance measures) and the factors that influenced the construction and application of a consistent procedure are examined. In the third section, the detailed performance measures are analyzed, including the number of problems answered correctly and solution time as a function of information and type of problem. In the fourth section, the relationship of the other four dependent measures to performance on the Computer test is discussed. This includes performance on the Procedure Protocols, Transfer test performance, Questionnaire responses, and Inductive test performance. Throughout the results section, the implications of the results for the model outlined in the introduction are highlighted, although a more elaborate treatment has been reserved for the discussion.

Part 1: Patterns of Information Use

Total time. An analysis of the total time spent on the instruction or example information was done with two between-subjects factors, Form nested within Information (Instructions, Examples). The Information effect was not significant, $F(1,68)<1.0$, although the Form effect was, $F(2,68)=3.53$, $p<.05$. Less time was spent examining the specific examples than the prose examples, $F(1,68)=7.05$, $p<.05$. The initial amount of time spent on reading the instructions or examples was also analysed. There was no effect of Information, $F(1,68)<1.0$, and the Form effect was not significant, $F(2,68)=1.74$, $p<.20$. These two Information time effects are shown in Figure 1.

The four types of Information were considerably different in length, as indexed by the number of words in each. Therefore, the lack of a significant initial or total time difference is somewhat surprising. If the time measure only measured the time necessary to read the presented information, it might have been expected that at least the initial times would reflect these differences in length.

Figure 1: Patterns of Information Use



However, the amount of time spent on the information may instead reflect the way in which subjects used that information.

Two supplementary analyses were done to examine this possibility. First, the initial time measure was divided by the number of words in each information source, 64 for the schematic instructions, 63 for the procedural instructions, 30 for the prose example, and 18 for the specific example. Then, this measure of time per word was analysed with Information (Instructions, Examples) and Form. The results of this analysis indicated that the Example subjects spent 1.06 seconds on each word, which is significantly longer than the 0.39 seconds Instruction subjects spent on each word, $F(1,68)=37.62$, $p<.001$. The effect of Form was not significant, $F(2,68)<1.0$. The example subjects were not just slower readers, since the median reading speed measure was approximately equal across conditions (described below). Instruction subjects' reading speed was about .30 second per word, and example subjects' reading speed was calculated to be about .34 seconds per word. The observed difference in time spent on the instruction and example information is consistent with the model: the time spent on the information reflects the use of different processes by example and instruction subjects.

The second supplementary analysis involved calculating the correlations between a measure of reading speed, and the amount of time subjects spent initially, and in total, on the instruction or example information. The correlations are shown in Table 6. The reading speed measure used was the median time calculated from the five initial screens of information that subjects read before they started the test. This reading speed measure was assumed to reflect a subject's normal reading speed. As is evident from the table, only the Schematic Instruction times (inital and total) were significantly correlated with reading speed, $ts(16)>3.40$, $ps<.01$. That is, the variability in time (initial and total) for Schematic Instruction users was at least partially accounted for by those subject's overall reading speed. The results of these two supplementary analyses suggest that the time spent on the instruction or example information reflects the operation of different processes. Example subjects may be

Table 6

Correlations Between Median Reading Speed and Time Spent on Instructions or Example

	Initial Time	Total Time
Type of Information:		
Instructions		
Schematic	.66[1]	.65[1]
Procedural	.28	.46[2]
Examples		
Prose	.45[2]	.27
Specific	.15	.03

[1]$p < .01$

[2]$p < .10$

beginning the construction of their procedures, whereas instruction subjects are attempting to retrieve the procedure that is indicated. According to the correlations shown in Table 6, reading speed seems to be more closely related to the processes involved in the three prose forms of the different types of information. Reading speed seems to be most closely related to the process of retrieving a labelled procedure (Schematic Instructions). Thus, despite a lack of difference in overall time spent on instructions and examples, there is evidence that the time is used differently.

Frequency of references. The total number of references made to the presented information by each subject was also analysed with a 2 by 2 design, with Form nested within Information. The Form effect was not significant, $F(2,68)<1.0$. However, the mean number of references made to the instructions was 2.89 which is significantly less than the mean of 3.86 references made to examples, $F(1,68)=5.54$, $p<.05$. The mean number of references made to each of the four information types is shown in the top part of Figure 1. This result is consistent with the model proposed in the introduction. Examples may be referred to more often than instructions during the operation of the analogical comparison process. Information is extracted from instructions and used to retrieve a procedure.

Pattern of references back. Most of the subjects in the 4 conditions looked back at the instructional information at least once: 14, 15, 16, and 16 subjects out of 18 respectively across the four conditions. However, the pattern of lookbacks is different as a function of Information. Seventy-five percent of the Example subjects' first references back occurred as they were solving or just after they finished solving the first problem, as compared to 45% of the Instruction subjects' first references back ($p<.01$, joint-binomial test). Instruction users tended to look back later in the course of the test. This result supports the hypothesis that example subjects use the example in a comparative way, especially at first. Example subjects refer back in the process of constructing and generalizing their procedure. Instruction subjects may only refer back to the instructions if their retrieved procedure becomes inadequate for a particular problem.

Summary. The total amount of time spent on instructions and examples is superficially similar until the lengths of the different forms of Information and the number of references are taken into account. Instruction use seems to be more closely related to reading speed than example use is, suggesting that the different types of information are used in conjunction with different processes. Furthermore, examples are referred to more often than instructions are, and the pattern of references to the example and instruction information are different. These results are consistent with the hypothesis that the example is used in conjunction with an analogical mechanism that may operate by comparing the example to the test problems and extracting the similarities.

Part 2: Consistent Procedure Use

In the first section, it was demonstrated that instruction and example use differs with respect to time, and the number and pattern of references. The usefulness of each type of Information, and performance across the whole test as a function of Information, are explored in this section.

Instructions. The percentage of subjects in each condition who used a consistent procedure is shown in Table 7. Consistent procedure use for a subject was determined by a two-tailed sign test with alpha equal to .05, the null hypothesis being that subjects randomly selected one of either the series or categorization response for each item. The series distractor was rarely chosen. Therefore, for this 36-item test, a consistent subject answered 23 or more questions with the series response.

For the Schematic Instructions, most of the subjects were able to use the information to consistently perform the correct procedure. However, this was not true for the Procedural Instructions. More subjects in the Procedural Instructions condition used a categorization procedure (instead of the instructed series procedure) than in the Schematic Instructions condition ($p = .046$, joint-binomial test). These differences between the two conditions suggest that the Procedural Instructions were more difficult to interpret than the Schematic Instructions. This is somewhat surprising, since the Schematic Instructions are quite vague and the Procedural

Table 7

Percentages of Subjects Using Each Type of Procedure as a Function of Information

	Type of Procedure		
	Series	Category	Undecided
Instructions			
Schematic	83	6	11
Procedural	56	28	17
Examples			
Prose	72	28	0
Specific	78	6	17

Instructions are more specific. This "instruction" in the Schematic case can be reduced to a single word (series), whereas the Procedural Instructions explicitly specify the steps in a procedure. Therefore, the major difference between the two conditions was that the Schematic Instructions provided the label of a stored, generalized or schematic procedure, whereas the Procedural Instructions described the steps in a possible procedure, without ever specifying a potentially retrievable label. According to the model of instruction use proposed by Dixon (in press), it may be difficult for subjects in the Procedural Instructions condition to retrieve a stored procedure without the higher level label, even if they possessed such a procedure. Therefore, a procedure would have to be constructed from the instruction information and this operation is potentially quite difficult, as is reflected by these results (Anderson et al., 1983).

The categorization procedure seemed to be a sort of 'default' procedure. Out of the three answer alternatives provided for each question, the category name can always be related to all of the problem words, although it does not complete the series. If a subject was unsure about the 'series' procedure, or did not retrieve a suitable procedure, the categorization procedure is simple and quite fast. In the protocols, a number of subjects also mentioned that they chose the category alternative when none of the other alternatives seemed to be correct. Therefore, the use of this procedure may signal an inability to retrieve a procedure.

The fact that more subjects in the Procedural Instructions condition reverted to use of the easier categorization strategy than in the Schematic Instruction condition is consistent with the idea that they had trouble retrieving a procedure. The series procedure is neither inherently difficult, nor obscure, since most subjects who were told to use a series procedure were able to do so. Therefore, this result supports the contention that provision of the appropriate level action in the hierarchy will be the crucial aspect of "good" instructions. The usefulness of step-by-step or detailed information such as that provided by the Procedural Instructions may be limited by the accessibility of the higher level information that assists proper interpretation (Smith & Goodman, 1982). The primary purpose of the

type of instruction information used in this study may be to activate knowledge structures that will guide subsequent performance, rather than to instruct in the sense of teaching a new skill or procedure.

Examples. A majority of subjects in the two example conditions were able to use the example information to perform consistently. However, more of the Prose Example subjects used a consistent categorization procedure than did Specific Example subjects ($p=.046$, joint binomial test), suggesting that it may be somewhat more difficult to construct a series procedure from the Prose Examples. The Prose Examples were intended to provide the same information that a subject might extract from a Specific Example. However, there were two potentially important differences between the Example types. The Prose Example did not illustrate a particular problem, and it did not include the answer alternatives that were wrong (counter-information). The relative difficulty of the prose example suggests that some subjects might find that examples presented in a prose format are difficult to interpret. The reason more subjects used the categorization procedure may have been related to the fact that there is no counter-information in the Prose Example that indicates the categorization strategy is wrong. Therefore confused subjects might be more likely to capitulate and use the easier procedure, since there was no direct evidence that it was incorrect.

Another interesting result involves the class of example and the consistency of procedure use. A breakdown of procedure use by the class of example provided is shown in Table 8. Ordinal examples were more useful than both Measure examples ($p=.0099$, joint-binomial test), and Animal examples ($p=.0166$, joint-binomial test). Furthermore, all subjects who were given Ordinal examples used a consistent series procedure. This may be because the 'series' aspect of the Ordinal problems is very obvious, and the ordinal series used were all very familiar. A good example therefore may be one that is most similar to or representative of all the problems, rather than the one that demonstrates how to solve the most difficult problems.

Summary. Consistent performance on the computer test varied as a function of the type of information presented to the subject. For instructions, providing the

Table 8

The Percentage of Example Subjects Who Used a Consistent Series Procedure

	Type of Example		
	Prose	Specific	Total
Class of Example			
Animal	50	83	67
Measure	67	50	58
Ordinal	100	100	100

steps in a procedure was less helpful than merely mentioning the name of a useful and well-known procedure. The assumption that most people would have a "series" procedure to retrieve that could be used seems to have been supported. Examples are just as useful as instructions in an absolute sense, although some types of examples may be more useful than other types. Overall, there is little evidence for an instruction or example advantage, but there was some indication that instructions are used to retrieve a procedure, whereas examples can be used to construct a procedure.

Part 3: Detailed Performance Measures

Three detailed performance measures were assessed. They included the number of problems solved correctly (accuracy), the amount of time spent on correctly solved problems (solution time), and performance on these two measures across the two halves of the test (practice effects), all as a function of the four problem types (Animal, Measure, Ordinal, and Other). In the Introduction, two predictions were made about these measures. First, a similarity effect was predicted for examples, such that problems similar to the example should be solved more quickly and more accurately than problems that were not similar to the example. The similarity effect was hypothesised to be a byproduct of the analogical mechanism. The second hypothesis was that practice effects should be greater for example users, since instruction users should have more consistent and applicable procedures initially, whereas example users are involved in the construction of a procedure.

Only those subjects who were able to interpret the instructions or examples successfully were included in these analyses, e.g., those who used a consistent series procedure. This included 15 Schematic Instruction users, 10 Procedural Instruction users, 14 Specific Example users, and 13 Prose Example users. The undecided subjects' and category users' data would not reflect performance as a function of Information since they were unable to use that information to perform consistently. In fact, it would be misleading to include the data from these subjects. The category strategy is much faster to execute than the series strategy, and so the category users' overall solution time would be less. Furthermore, if a series response

is considered to be correct, some category users have only a few or zero "correct" responses, since they chose the category response. Including category-procedure users data in the accuracy analyses would be unrepresentative, since they actually chose the correct response according to the procedure they were using, although the procedure they chose was inappropriate given the Instructions or Example.

Solution time. Solution time (the median of each subject's times for correct answers for each Problem Type), was analyzed in a three-way analysis of variance, with two between-subjects measures, Form within Information, and one within-subjects measure, Problem Type (Animal, Measure, Ordinal and Other). There was no effect of Information or of Form on solution time, F's<1.0. However, the effect of Problem Class was highly significant, $F(3,144)=10.44$, $p<.001$. The mean solution times for each problem type are shown in Table 9.

Animal problems were solved significantly more slowly than all other types of problems, ($p<.05$, Tukey HSD). Animal problems may be solved most slowly because the series dimension is not necessarily familar to all subjects. People may not be used to comparing different animals on a size dimension and then making judgments about them. The Measure and Ordinal problems are more familar as series, and the dimensions on which they varied may have been more obvious. The Other series may have had only one obvious dimension on which to vary which would reduce solution time.

Accuracy. The identical analysis as described above was done on the accuracy measure (the number correct in each problem class for each subject). There were again no effects of Information, $F(1,48)<1.0$, or of Form, $F(2,48)=1.34$, $p=.271$. However, there was an effect of Problem Class, $F(3,144)=28.59$, $p<.001$. Animal problems were solved significantly less accurately than all others, ($p<.05$, Tukey HSD test). The mean accuracy values are shown in Table 9.

Similarity effects. The similarity effect was explored by combining the two example conditions (in order to increase the sample size), and then doing two-way analyses of variance on solution times and accuracy. The between-subjects factor was the Class of Example (Animal, Measure, Ordinal) that the subject was given,

Table 9

Mean Solution Time and Mean Accuracy Values for the Four Problem Types

	Response Time (seconds)	Accuracy (percent)
Type of Problem		
Animal	11.6 (4.74)	72 (24.7)
Measure	9.4 (4.24)	89 (10.8)
Ordinal	9.6 (2.89)	94 (9.9)
Other	8.4 (2.87)	91 (11.8)

Note: Standard deviations appear in parentheses.

and the within-subjects measure was Problem Type (Animal, Measure, Ordinal). For any subject then, one of the three Problem types was similar to the example that they were given, and the other two problem types were dissimilar to the example.

For the response time measure, there was an effect of Class of Example, such that the Measure example subjects were fastest and the Ordinal example subjects slowest, $F(2,24)=3.61$, $p<.05$. The Measure example category also had the fewest subjects in it, so this result could reflect a selection process, such that the slower subjects are also the ones who could not use the Measure example and so are not included in the analysis. Since all the Ordinal example subjects are included in the analysis, no such selection of faster subjects would be taking place. The effect of Problem Type was also significant, $F(2,48)=4.35$, $p<.05$ and as before, Animal problems were solved most slowly. There was no similarity effect for times. The Class of Example a subject received did not result in an advantage for problems that were similar to the example. The mean solution time values are shown in Table 10. A similarity effect would be indicated if the values on the left to right diagonal were smaller than others in the same row. That is, subjects who got an Animal example would be faster on Animal problems than the subjects who got other types of examples.

Accuracy was also analyzed for the example subjects with Class of Example (Animal, Measure, Ordinal) and Problem Type (Animal, Measure, Ordinal). There was no effect of the Class of Example, $F(2,24)<1.0$. But, as in the above analysis, there was an effect of Problem Type on accuracy, $F(2,48)=12.43$, $p<.001$. Animal problems were solved less accurately than the other types. A planned comparison contrasting problems similar to the example versus dissimilar indicated that the similar problems were solved more accurately, $F(1,48)=4.13$, $p<.05$. This replicates and extends the similarity effect that was suggested by the data collected by LeFevre and Dixon (1984): even when subjects are consistent across the test, they are still more accurate on problems that are similar to the example. The similarity effect for each problem type is shown in Table 10. The largest effect was for the animal problems and the smallest for the ordinal problems, that is, the

Table 10

Mean Solution Time and Mean Accuracy Values for the Similarity Analyses

	Type of Example Subject was Given		
	Animal	Measure	Ordinal
Mean Solution Time (in seconds)			
Animal	11.1 (3.8)	9.5 (3.6)	13.6 (6.3)
Measure	10.6 (4.0)	7.0 (1.8)	9.9 (3.4)
Ordinal	9.4 (2.2)	7.5 (1.9)	10.9 (2.2)
Accuracy (percent correct)			
Animal	82 (8.3)	78 (19.2)	67 (28.8)
Measure	86 (11.5)	92 (12.4)	90 (11.1)
Ordinal	90 (19.2)	95 (8.7)	96 (9.8)

Note: Standard deviations appear in parentheses.

similarity effect appears to be greatest for the most difficult type of problem. However, the other two problem types may be subject to a ceiling effect.

Practice effects: Solution time. The hypothesis outlined in the Introduction with respect to practice effects was that the example subjects should show larger practice effects than instruction subjects. To test this hypothesis, a four-way analysis of variance was done on the median solution times for each subject, contrasting performance on the first 16 problems to performance on the last 16 problems. Originally, these had been designed to be balanced across subjects and problems; however, the data from a substantial proportion of the subjects are missing because these subjects failed to use a consistent series procedure. Therefore, the balancing is incomplete. Subjects received four of each problem type in each half of the test. The between-subjects factors were Form within Information and the two within-subjects factors were Practice (first half, second half) and Problem Type (Animal, Measure, Ordinal, Other).

For the solution time measure, there were a relatively large number of subjects who had missing data points for the animal problems in the first half (8 subjects altogether out of 52 and 4 in the Schematic Instruction condition alone). For the first analysis discussed, these missing points were replaced by the mean across problem types for that subject. This procedure would bias against getting either a Practice or Problem Type effect. For this first analysis, there were no effects of Information, $F(1,48)=1.44$, $p=.236$, or of Form, $F(2,48)=1.15$, $p=.326$. However, there was a significant Practice effect, $F(1,48)=16.92$, $p<.001$, such that subjects were faster by about 2 seconds on the second half of the test. There was also a significant effect of Problem Type, $F(3,144)=9.71$, $p<.001$. The animal problems were again solved most slowly. The relevant means (Table 11) indicate that overall, subjects get faster with practice.

In the second analysis, the animal problems were not included because of the potential problems with missing data. As before, the data were analyzed with three factors, Information, Practice and Problem Type (Measure, Ordinal and Other). There were no effects of Information, $F(1,48)=1.06$, or of Form, $F(2,48)=1.0$, or

Table 11

Mean Solution Time as a Function of Practice and Information Type

	First Half	Second Half
Information Type		
Instructions		
Schematic	9.7 (3.6)	9.4 (4.4)
Procedural	12.6 (7.6)	9.6 (4.2)
Examples		
Specific	11.8 (7.4)	9.5 (4.2)
Prose	13.7 (5.6)	10.4 (4.3)

Note: Standard deviations appear in parentheses.

of Problem Type, $F(2,96)<1.0$. The only significant effect in this analysis was that of Practice, $F(1,48)=15.64$, $p<.001$.

In a subsidiary analysis for practice effects, Schematic Instruction subjects' performance was compared to that of the Specific Example subjects, since these conditions are postulated to be least like each other in terms of information use. The Schematic Instruction condition is a "pure" case of specifying a retrievable label, and the Specific Example condition involves the type of example most amenable to a direct comparison process. The analysis was done with three factors, Information (Schematic, Specific), Practice (first half, second half) and Problem Type (Animal, Measure, Ordinal and Other). There was no effect of Information, $F(1,27)=1.34$, $p=.26$. The Practice effect was smaller for this comparison than in the overall analysis, $F(1,27)=7.62$, $p<.02$, but the animal problems were still solved most slowly, $F(3,81)=5.8$, $p<.002$. The interaction of Practice with Information was marginally significant, $F(1,27)=3.61$, $p=.068$. This effect is shown in Figure 2. It is consistent with the hypothesis that instruction users improve less than example users. According to the proposed model, example users are initially involved in procedure construction, whereas instruction users are simply executing retrieved procedures throughout the course of the test. By the second half of the test, example users also seem to be primarily involved in procedure execution.

Practice effects: Accuracy. Accuracy, or the number of correct series responses, was also analyzed in a 2 by 2 by 2 by 4 design. The between-subjects factors were Form within Information, and the within-subjects factors were Practice and Problem Type. There were no main effects of Information, $F(1,48)<1.0$ or Form, $F(2,48)=1.29$, $p=.285$. As with the time measure, there was a significant practice effect, $F(1,48)=7.41$, $p<.01$, such than the overall accuracy increased from 83% in the first half to 89% in the second half. There was also a significant effect of Problem Type, $F(3,144)=26.39$, $p<.001$, Animal problems being least accurately solved. Furthermore, Practice interacted with Problem Type, $F(3,144)=7.91$, $p<.001$. As can be seen in Figure 3, the Practice effect was specific to the Animal problems. They showed substantial improvement from the

Figure 2: Practice Effects for Schematic Instructions and Specific Example Conditions

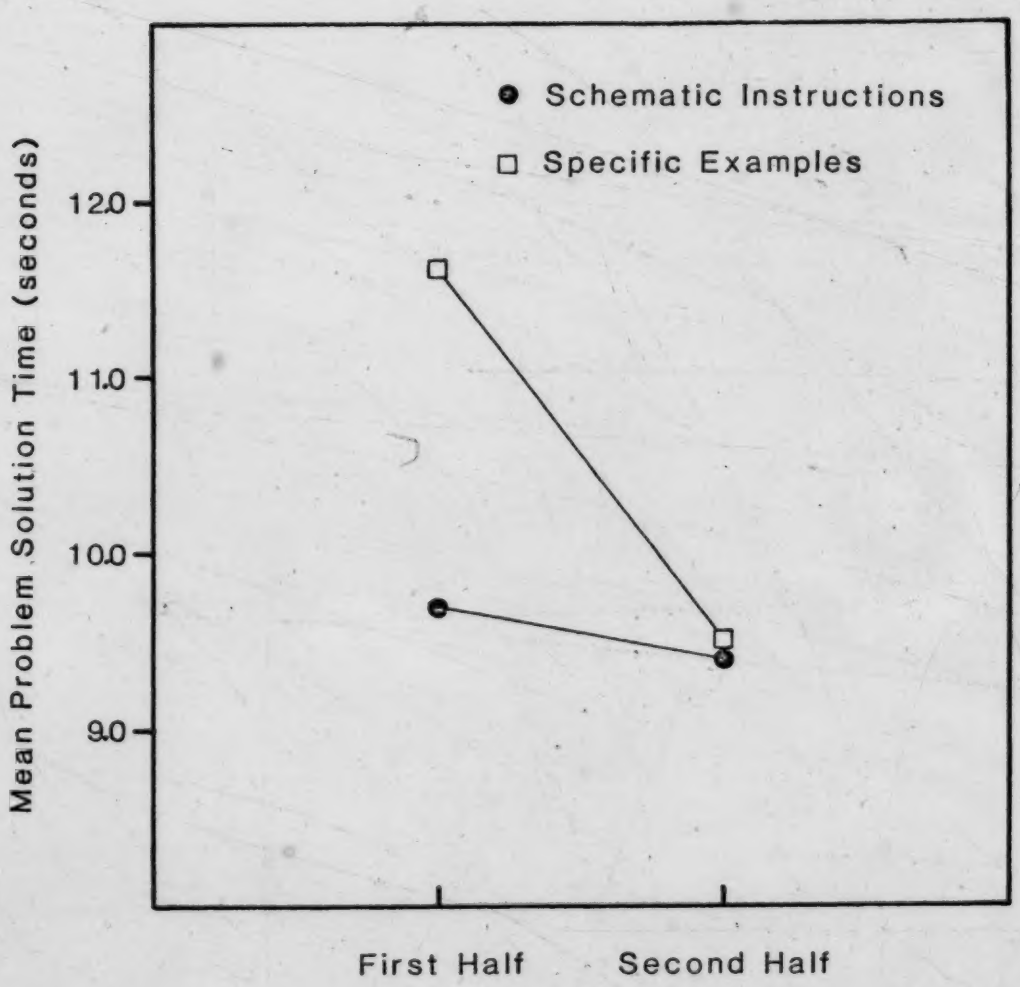
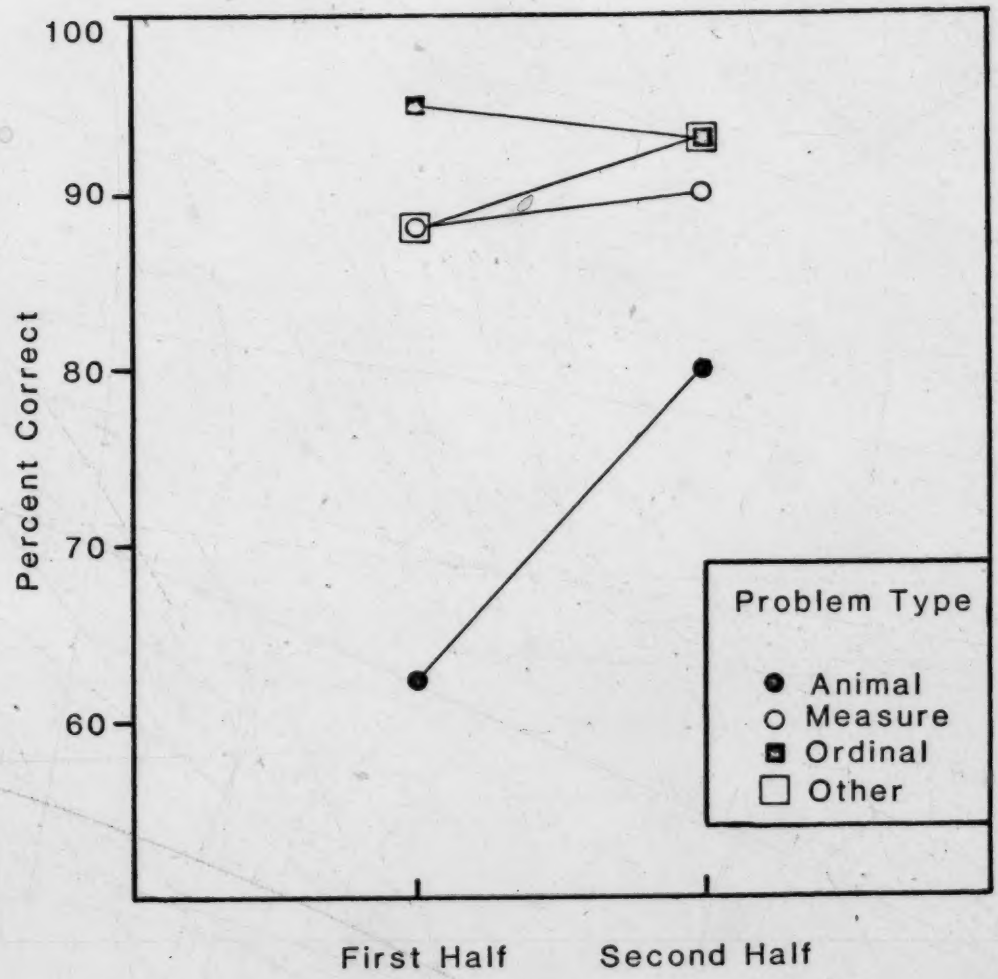


Figure 3: Practice Effects for Accuracy by Problem Type.



first to the second half. The other problem types may have been constrained by a ceiling effect.

There was also a slight interaction of Information and Practice, $F(1,48)=2.99$, $p=.09$ (see Figure 4). The accuracy of the example subjects improved more than that of the instruction subjects. As is evident from Table 12, the differential improvement for example subjects is consistent across Forms. Further, the variability in percent correct decreases across halves for example subjects although it does not decrease for instruction subjects. The example subjects seem to be improving and becoming more stable in their performance, as was suggested by the model. However, since accuracy seems to be constrained by a ceiling effect and the interaction is only marginal, the hypothesis presented in the Introduction is only weakly supported by the practice effect data.

Summary. Support for the two major predictions of similarity and differential practice effects was reviewed in this section. According to the similarity hypothesis, an example should facilitate those problems from the same class. This was supported in the accuracy data, but not in the solution time data. The major prediction derived from the practice effect hypothesis was that instruction subjects would not improve much with practice relative to example subjects. This hypothesis was weakly supported by selected portions of the solution time and accuracy analyses. In sum, although these effects may be present, they were not strongly demonstrated with these data. Limitations of this research for testing the similarity and practice effect hypotheses and reasons for the lack of support are presented in the Discussion. It should be noted that the degree to which the Transfer test is actually related to the Computer Test is not clear. In some ways, the two tasks are similar since they were created by the same person for the same purpose. Both have two essentially correct procedures that can be used. However, although one of these procedures is "series" for both tasks, the other two procedures are different. For the series-classification test, using a procedure entails ignoring some features of the problem and concentrating on others. In some ways, this conflicts with the procedure necessary for the Computer Test. In sum, the two tasks may be

Figure 4: Practice Effects for Accuracy by Information

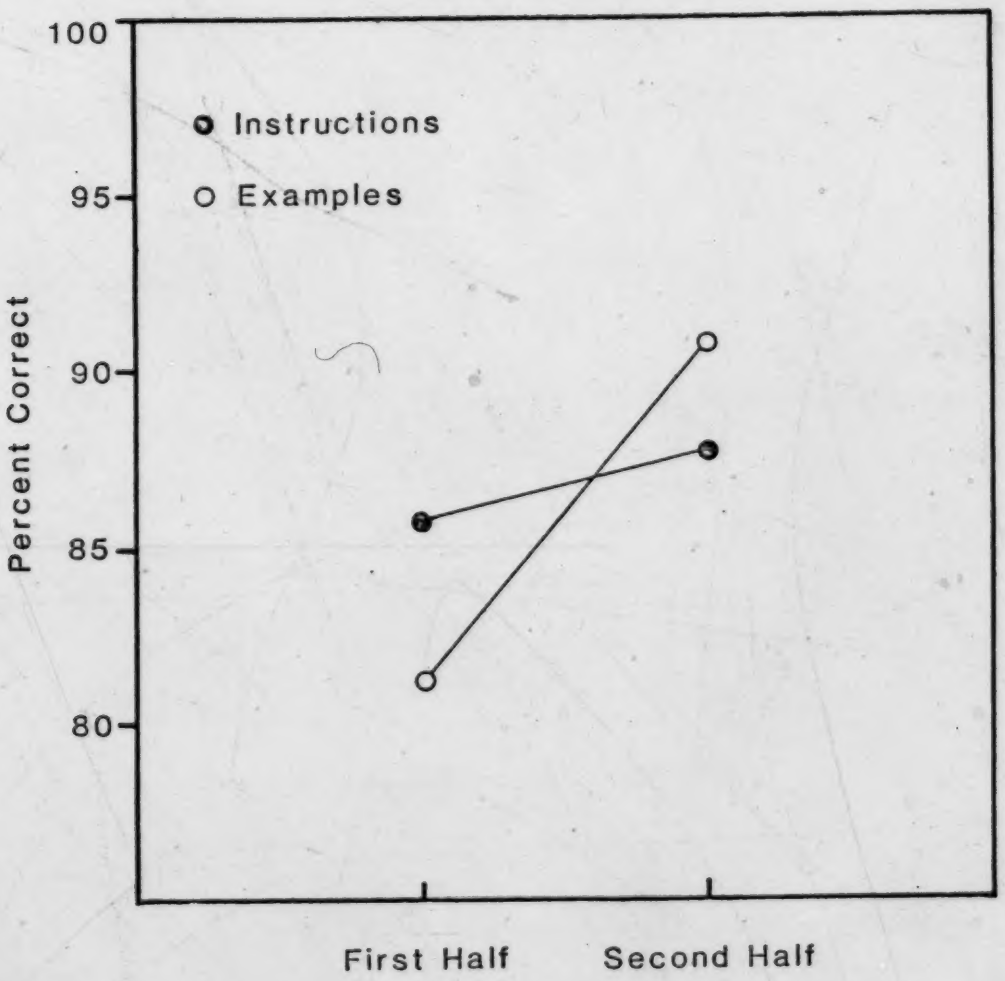


Table 12

Mean Accuracy as a Function of Practice and Information Type (percent correct)

	First Half	Second Half
Instructions		
Schematic	83 (18.7)	87 (18.3)
Procedural	89 (16.0)	89 (21.4)
Examples		
Specific	85 (18.8)	92 (14.4)
Prose	79 (27.0)	89 (15.0)

Note: Standard deviations appear in parentheses.

somewhat less similar than they appear.

Part 4: Other Dependent Measures

Performance on the other four dependent measures -- the Procedure Protocol Description, the Transfer test, the Questionnaire, and the Inductive-Reasoning test -was analyzed with attention to two factors, the type of Information the subject received, and whether or not they were able to interpret that information and perform consistently (Procedure Use).

Procedure protocol description. Subjects were asked to write a short description of how they solved the problems on the test. These protocols were scored on a number of measures by two independent judges. The judges were blind to the type of Information that a subject received. Each measure and its relevance will be discussed in turn. On the first three measures, all of which involve the presence or absence of some feature, agreement between the judges was quite good. On the judgmental measures (goodness and generality) agreement was also fairly good, and the trends were the same for both judges' ratings. The motivation for asking subjects to describe their procedure involved the question of whether this sort of description would reflect any of the predicted differences in generality of example- versus instruction-generated procedures.

1. Was there evidence that subjects used a series procedure?

The protocols were scored according to whether it appeared that the subject had used a series procedure, a categorization procedure, or it was not possible to discriminate. Comparison of actual procedure use to the decision of the judges is shown in Table 13. Agreement between judges was 89%, and it seems that this measure was fairly close to the actual use of series procedures. Most of the subjects who used a series procedure were able to clearly indicate this in their written protocols. This result is important, since they could all have been using some other procedure and by chance have given the series answer.

Table 13

Comparison of Actual Procedure Use to Judged Procedure Use

	Instructions		Examples	
	Schematic	Procedural	Prose	Specific
Used a Series Procedure				
Consistently	15	10	13	14
Judge 1	17	11	14	14
Judge 2	17	11	14	13
Used a Category Procedure				
Consistently	0	5	5	3
Judge 1	0	4	2	0
Judge 2	0	4	2	1
Unclassifiable				
Consistently	3	3	0	1
Judge 1	1	3	2	4
Judge 2	1	3	2	4

2. Were Specific Examples given in the protocol?

A specific example was a problem that resembled those on the computer test, with or without all of the possible answer alternatives included. Four Instruction subjects (2 in each condition) and ten of the example subjects (6 in the Prose and 4 in the Specific conditions) included specific examples in their protocols. The example subjects were more likely to include specific examples ($p=.032$, joint-binomial test). There are two different explanations for this effect. Either subjects remembered the nature of the initial information and tried to imitate it when asked to describe their procedure, or because of the nature of their procedures, example subjects relied on examples more. Subjects tended to give examples that came later in the test: of the 9 that were unequivocably comparable to actual test problems, the median position of the relevant problem was 28 out of 36. They did not reproduce the particular example that they were given.

In total, 18 specific examples were given by the subjects. Some subjects gave more than one example. Of these 18 examples, 10 were Ordinal Class examples, 4 were from the Animal class, 1 from Measure and 3 from the Other class. Furthermore, 10 out of the 14 subjects gave at least one Ordinal example. This preponderance of Ordinal examples supports the idea that Ordinal problems are most representative of all the problems, or perhaps most representative of the series procedure.

The examples given in the protocols varied in quality. An example was either (a) an exactly reproduced test question, (b) a potential test question that had not actually appeared, or (c) wrong, in that it could not have appeared on the test. The quality of the examples given by the subjects is shown in Table 14. All the given examples used words that had been given on the test, but only 33% of them were actual test problems. It is interesting to note that subjects who used the categorization procedure and provided an example often did not preserve the "series" present in the first three words. For instance, one subject gave as an example "Monday, Tuesday, Saturday", indicating that "days" was the correct answer. The actual problem was "Monday, Wednesday, Friday", suggesting that the subject had

Table 14

Frequency of Specific Examples Given in Procedure Protocols

	Instruction Subjects	Example Subjects	Total
Quality			
Same as Test Problem	1	5	6
Not Same, but Correct	1	5	6
Wrong	2	4	6

not noticed or remembered the series present in the problem words. To these subjects, the series aspect of the problems was irrelevant to solution (since they chose the category alternative) and so presumably was not encoded, or at least not included in their representation of the problems. These are the 'Wrong' examples in Table 14.

3. Did they mention using an analogy strategy?

An analogy strategy involves comparing the first two words in the problem, finding the similarity between them, then choosing the answer alternative that represents this same relationship with the third problem word. This procedure will produce the same answer as the series procedure produces. In the Instruction conditions, one Schematic subject said that he or she used an analogy strategy. In the Example conditions, one Prose example and four Specific example subjects mentioned that they used an analogy procedure at some point during the course of the test. The Specific example subjects showed a slightly larger tendency to use this procedure. This could be due to the fact that the analogy procedure is not specifically contradicted by the specific example as it might be by the other conditions. This result suggests that examples leave open more of a possibility that subjects will construct a personal or idiosyncratic procedure. For this task, it was not crucial although it might be an important consideration for other tasks.

4. Generality

Each protocol was rated on a five point scale, where 5 indicated a completely general or schematic protocol, and 1 was a completely specific description. The correlation between the two judges ratings of the protocols was .81 ($p<.01$), and there was 94% agreement within one point of the scale values. Examples of protocols are given in Table 15. Generality ratings were analyzed in an analysis of variance with three between-subjects measures, Form within Information by Procedure Use. There was a significant effect of Information, $F(1,64)=6.63$, $p<.02$. The mean generality rating for examples was 3.3, and for instructions was 4.1. There are two interpretations of this result. Either subjects remember the initial information they are given and try to reproduce it, or the instruction- or

Table 15

Examples of Protocols

Rating: Generality (5), Goodness (1)
First try to figure out the relationship between the words. Then follow the same line of thought and choose the answer.
Rating: Generality (2), Goodness (4)
Three words were presented in some order (eg. lowest to highest or smallest to largest). Based on the example given I assumed the researcher wanted the next word in the sequence instead of stating how they were common or the item that could have appeared at the beginning of the list.

example-generated procedures are actually different in specificity. The latter interpretation is consistent with the proposed model of instruction and example use. The three-way interaction of Form within Information by Procedure Use was also significant, $F(2,64)=4.40$, $p<.03$. The mean generality ratings are shown in Table 16. This interaction is probably due to the fact that the protocols of nonseries users in the prose example condition were quite specific, that is, the ratings were quite low.

5. Goodness.

The protocols were scored on a five point scale for "goodness", where 5 meant that the protocol was extremely clear and comprehensible, and 1 meant that it was totally incomprehensible. The criteria for rating included: How clear and understandable is the description? Could the rater tell what the subject was doing? Would someone else be able to use this description to do the test? The correlation between the two judges ratings was .85 ($p<.01$) and there was 100% agreement within one point of the scale. The ratings were analyzed in a three-way analysis with Form within Information by Procedure Use, as above. There were no significant effects of Information or of Form (F's <1.0) or of Procedure Use ($F(1,64)=2.40$, $p=.126$). However, the interaction of Information and Procedure Use was significant, $F(1,64)=6.93$, $p<.02$. This interaction is shown in Figure 5. For instructions, the quality of the protocol did not seem to be systematically related to performance. Some of the Procedural Instructions subjects gave good categorization descriptions; others may have been trying to reproduce what they remembered from their instructions, in spite of how they performed. For example users however, whether or not they were able to interpret the example was more important to the quality of the description. People who were able to use the example tended to give better descriptions, perhaps because the non-example users had no memory of instruction information to fall back on.

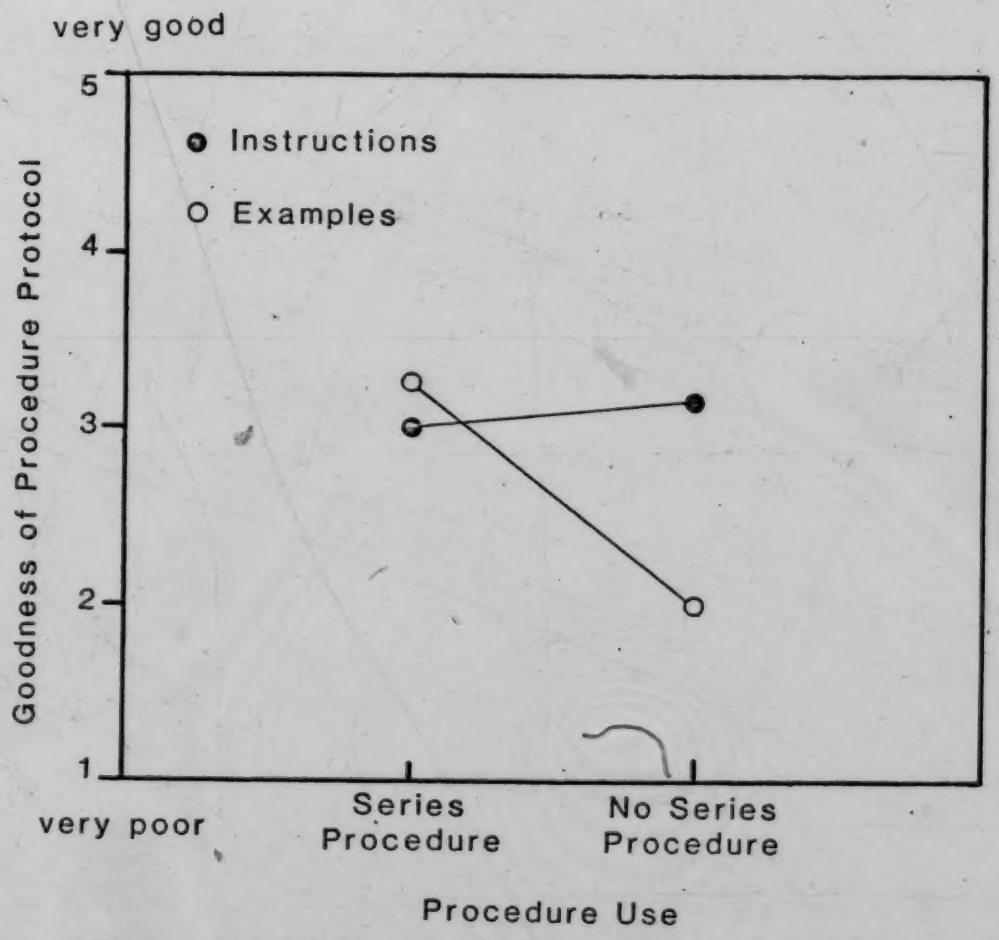
Table 16

Length, Generality and Goodness Measure Means as a Function of Information Type and Procedure Use

	Instructions		Examples	
	Schematic	Procedural	Prose	Specific
Length				
Series	43 (16.0)	48 (27.9)	48 (22.9)	51 (13.3)
Not series	38 (11.5)	43 (23.0)	38 (18.7)	34 (25.3)
Generality				
Series	3.9 (1.2)	3.9 (1.1)	3.4 (1.5)	3.3 (1.4)
Not series	4.0 (1.0)	4.8 (0.7)	4.8 (1.1)	2.0 (0.5)
Goodness				
Series	3.3 (0.8)	2.6 (1.4)	3.4 (1.2)	3.1 (0.9)
Not series	3.0 (1.0)	3.2 (1.3)	1.8 (0.9)	2.2 (0.5)

Note: Standard deviations appear in parentheses.

Figure 5: Mean Goodness Rating as a Function of Procedure Use



6. Length (Number of Words)

The number of words in each protocol was calculated and analyzed. It is possible that judgments of goodness and generality could simply be a result of length. Therefore, length was analyzed as a function of Information and correlated to the other two measures.

The length of the protocols as indexed by number of words ranged from 16 to 100 across subjects. The number of words was used as the dependent measure in a 2 by 2 by 2 analysis, with Procedure Use crossed with Form nested within Information. There were no significant effects of either Procedure Use, $F(1,64)=3.02$, $p<.10$, Information, $F(1,64)<1.0$, or Form, $F(2,64)<1.0$. The mean number of words for each condition are shown in Table 16.

7. Inter-relations between measures.

For both judges, generality and goodness were not significantly correlated. However, generality was significantly negatively correlated with length (-.21 and -.33, $p<.05$) and goodness was positively correlated with length (.62 and .55, $p<.01$). General descriptions tended to be shorter probably because specific information tends to take longer to explain, and good descriptions tended to be longer. This was opposite to the relative lengths of the initial information provided to the subjects.

In sum, the protocol results suggest that instruction subjects can articulate a more general or schematic procedure, but these are not necessarily better than more specific, example-generated procedures. Example subjects are more likely to include examples in their protocols, and these tend to be Ordinal examples. Instruction subjects tend to be able to produce descriptions that suggest that they used a general, retrieved procedure, and this is unrelated to their ability to use this procedure to perform consistently.

Transfer test performance. Performance on the Transfer test was analyzed in two ways. Initially, an analysis of variance was performed using the number of series responses on the Transfer test as the dependent measure (out of 15 possible), and three between-subjects factors, Form within Information by Procedure Use. There was no effect of Information, $F(1,64)=1.97$, $p=.16$ or of Form

$F(2,64)=1.59$, $p=.211$, but there was a significant effect of Procedure Use, $F(1,64)=14.88$, $p<.001$. Subjects who used a series procedure on the Computer test also gave more series responses on the Transfer test (a mean of 8.7 for the 52 series users and a mean of 3.8 for the 20 non-series users). There was a tendency for series users to transfer their procedure from the first test to the second, as instructed.

However, *some* transfer is not really indicative of the use of a consistent series procedure on the Transfer test. The criterion for a consistent series procedure on the Transfer Test is 12 out of 15 possible responses consistent with the series strategy (as determined by a two-tailed sign test with alpha equal to .05). This strict criterion was used to examine procedure transfer from the Computer to the Transfer test. Table 17 shows the percentage of series users on the Computer test who also performed consistently on the Transfer test. As can be seen from the table, this varied considerably between conditions. Significantly more Specific Example subjects transfered than Schematic Instructions subjects ($p=.025$, joint-binomial test), though none of the other comparisons were significant. This result is counter to the hypothesis presented in the Introduction where it was suggested that a more general instruction procedure would transfer more readily than an example procedure. One explanation of this result is that a procedure generated by a specific example is most useful for transfer to another, related task, and a retrieved procedure is least useful. This possibility is examined in the Discussion.

A simple explanation of the differences in Transfer performance is that the overall quality of the series procedure was better for the Specific Example subjects. To test this, the total number of series responses made by each subject who responded consistently on the Computer test were analyzed using Form within Information as the factors. There was no difference in the number of series responses across conditions, F's<1.0. There is no evidence then that specific example subjects had better procedures on the first test.

Questionnaire data. The purpose of the questionnaire was to try and probe subjects' intuitions about example and instruction information. There were five

Table 17

The Percentage of Consistent Computer Subjects Who Used a Consistent Series Procedure on the Series-Classification Test

Instructions	
Schematic	27
Procedural	40
Examples	
Prose	38
Specific	64

metacognitive questions on the questionaire that subjects filled out. All used five point scales where 5 was the "good" or "more" end, and 1 was the "poor" or "less" end. Each will be discussed in turn. Three-way analyses were done with the factors Form within Information by Procedure Use and reliable results are discussed for each question. A summary of the means is shown in Table 18. Since these data were collected from the same subjects, the tests done on each question may be correlated. Results that are not significant at $p<.01$ should probably not be taken too seriously, since the error rate may be inflated.

Question 1. How useful did you find the (instructions/example)?

There were no differences across Information conditions in how useful the subjects perceived the different types of instruction and example information to be, F's<1.0, p's$>.25$. All were considered moderately useful. The overall mean was 3.3 out of 5.

Question 2. How much did you use the (instructions/example)?

There were no effects of Information on how much subjects said they used the instruction or example information, but there was a significant effect of Procedure Use, $F(1,64)=7.32$, $p<.01$. Series users said that they used the instructions more than non-users did, 2.9 vs 1.9 on a five point scale. This is an unfortunate complication of studying instruction/example use. Some people pay no attention to the instructions/examples in spite of their form or how useful they are. Not surprisingly, these people do not perform according to the instruction/example information. Curiously, although there were no differences in how useful these two groups of people thought the information was (Question 1), some did not use it.

Question 3. How useful do you usually find instruction sets?

There were no significant effects, although the effect of Information approached significance, $F(1,64)=3.31$, $p=.073$. The instruction people said that they usually find the instruction sets more useful than example subjects, 3.5 vs 3.0 out of 5, although these numbers indicate that the information was only moderately useful for all subjects.

Table 18

Mean Values on the Metacognitive Questionnaire as a Function of Information and Procedure Use

	Question Number				
	1	2	3	4	5
Instructions					
Schematic					
Series	3.7	2.7	3.7	3.7	3.6
	(1.3)	(1.6)	(1.0)	(0.9)	(0.9)
Not	4.3	3.3	4.0	3.7	4.7
Series	(0.6)	(1.2)	(1.0)	(1.5)	(0.6)
Procedural					
Series	3.2	3.0	3.3	4.3	4.5
	(1.4)	(1.3)	(0.7)	(0.7)	(0.7)
Not	2.9	1.5	3.1	4.2	3.2
Series	(1.6)	(0.8)	(1.0)	(0.7)	(1.3)
Examples					
Specific					
Series	3.5	3.1	3.1	3.9	3.6
	(1.4)	(1.4)	(1.0)	(0.7)	(1.3)
Not	2.2	1.8	1.8	3.8	3.8
Series	(1.1)	(1.1)	(1.2)	(1.1)	(0.9)
Prose					
Series	3.1	2.8	3.4	3.8	3.8
	(1.2)	(1.5)	(1.2)	(0.6)	(1.0)
Not	3.2	1.8	3.0	3.6	3.6
Series	(1.5)	(1.5)	(1.0)	(1.0)	(1.9)

Question 4. How easy was this test for you?

All subjects thought that the test was quite easy, and there were no effects of the type of Information or whether they used a Series Procedure, F's<1.5, p's$>.30$. The mean was 3.9 out of 5.

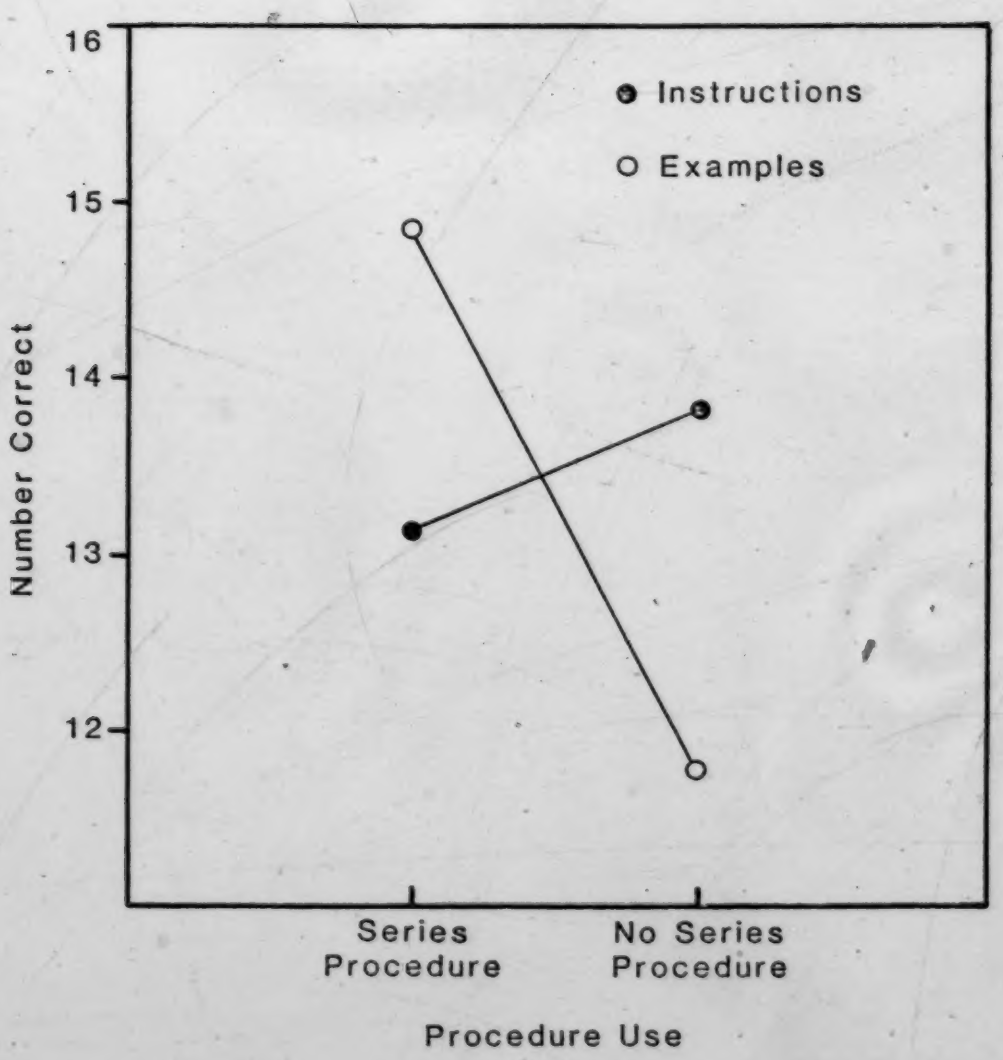
Question 5. How useful would (an example/instructions) have been? (That is, the opposite of what they were actually given).

This question was an attempt to again check if subjects always find examples more useful, or if experience with one type of information affects how useful they believe information to be. There were no main effects, F's<1.0. However, there was a significant interaction of Form, Information and Procedure Use, $F(2,64)=3.68$, $p<.03$. There were minimal differences for the example subjects between users and non-users; the means range from 3.6 to 3.9. All thought that instructions would have been moderately useful. However, schematic instruction non-users thought an example would have helped them a lot (4.7), whereas users thought that an example would have been only moderately useful (3.6). The Procedural Instruction subjects showed the opposite effect: non-users (most of whom used the category procedure) didn't think an example would be much help (3.2), but the users (who may have had to struggle with the interpretation of the instructions) thought an example would have been useful (4.5). In sum, the relative ease with which subjects interpreted the information seemed to have an important influence on whether or not they thought that it was useful.

The questionnaire data showed first that although subjects thought that all types of information were moderately useful, some tended to use the information more than others. The latter were also the people who did not use a consistent series procedure. Second, the difficulty of interpreting the instructions determined how important instruction subjects thought an example might be. Third, there was no evidence for an example bias among subjects who actually used one of the two information sources as found by LeFevre and Dixon (1984). The relative usefulness of Information and subject biases may depend on the task, how difficult the instructions or examples are to interpret, and the presence of conflicting information.

Number series test. Performance on the Number Series test was analyzed with the number correct as the dependent measure and three between-subject factors, Form within Information by Procedure Use. There were no significant main effects, F's<1.0, and the only significant interaction was between Information and Procedure Use, $F(1,64)=4.29$, $p<.05$. This interaction is illustrated in Figure 6. Subjects who successfully used the example scored higher on the number series test than did unsuccessful example users, whereas there was no difference on the number series score for series users and nonusers in the instruction conditions. This advantage for "good inducers" in the example conditions supports the idea that examples are used analogically since analogical reasoning is one form of induction. An alternate interpretation is that since the Number Series test also uses examples, this result indicates that example users do well on both tests. However, it is generally accepted by the test makers and those who use the tests to discriminate among people that IQ tests such as this are entrenched (Sternberg, 1981), and therefore that people already know how to do the problems. Thus, performance on an IQ test such as this one is assumed to vary with inductive reasoning ability, rather than with the ability to interpret examples.

Figure 6: Number Series Score by Information and Procedure Use



IV. Discussion

The main purpose of this thesis was to test a proposed model of differences in the way specific example and written instruction information are used in the process of constructing and using a procedure. A study was done in which four types of information (two types of instructions and two types of examples) were given for an inductive-reasoning test. Performance was observed to vary as a function of the type of information that was given. Some of the results of this study generated support for the model of instruction and example use that was proposed. To sum up the support for the model, the four hypotheses presented in the Introduction will be reviewed and the evidence that has been generated for them will be discussed. Then, implications of the results for the proposed model will be considered. Finally, new directions for research on instructions and examples will be proposed.

The first hypothesis presented in the Introduction was that problems that were similar to the example would be solved more quickly and with fewer errors than problems that were less similar to the example. The accuracy data supported this hypothesis: subjects were more accurate on problems that belonged to the same class as the example that they were given than on dissimilar classes. This was true even for subjects who were consistently using the correct procedure across the whole test, suggesting that it is a robust effect.

There are some important practical implications of the similarity effect. It suggests that the pattern of performance on a task (especially a more difficult one) could be partially a function of the example(s) that were given. This may be an important consideration for some tasks. Therefore, it may be best to provide as wide a range of examples as is feasible, or to try and give examples that are most representative of the optimal procedure for doing the task (Elio & Anderson, 1984). For instance, the most readily used type of example for the Computer task used in this experiment was from the Ordinal Class. Not only were all subjects who were given Ordinal examples able to use these to perform consistently, but most of the subjects who generated examples in the Protocol descriptions gave Ordinal examples.

The similarity effect did not hold for response time however: subjects were not consistently faster on problems that were similar to the example. There are a number of possible reasons for the failure to demonstrate a similarity effect with response time. First, both within-subject and between-subject variability on problems was quite high. This issue could be addressed by having a better model of the components of solution time, and measuring those components separately. As mentioned previously, the similarity effect may only be important for one part of the total problem solution time. For instance, deciding on the dimension of the change and selecting the answer may be a minor part in terms of the overall time spent on a problem, but it could be affected by similarity to a large extent. Encoding the words might constitute a larger and more variable segment of the total time, but it may be unaffected by similarity. The measure of response time that was used may not be sensitive enough to detect a similarity effect.

A second possibility is that a selection factor was operating that helped to reduce the response time similarity effect. Slow subjects may also be those who could not interpret the examples, and so were not included in the analyses. Furthermore, they may also be the subjects who are more likely to show a similarity effect. A third possibility is that similarity influences response time differently than it influences accuracy, although it is not obvious what this difference might be. Overall, it seems that a more precise model of the solution process may be necessary in order to be able to test an adequate similarity hypothesis for response time. Thus, the similarity effect obtained for accuracy supported the proposed model, although the results for the response time measure did not.

The second hypothesis that was described in the Introduction suggested that instruction subjects would show less of a practice effect than example subjects. There was only suggestive evidence from the Practice Effects analyses that this was the case. The effect was marginally significant for accuracy and for the response times of Schematic instruction versus Specific Example subjects. This suggests that

some kind of practice effect difference is present, although the data do not allow a firm acceptance of the hypothesis. Again, there are a number of reasons that this hypothesis was not adequately tested. First, accuracy measures were subject to ceiling effects, and time measures may have been subject to variability and floor effects. It could be that the practice difference would be more easily obtained on difficult tests where there is more room for substantive improvement. Second, the important practice effect might have been localized in one particular portion of the solution time component (as for the similarity effect). A better understanding of the problem solution process may be necessary in order to satisfactorily decompose the solution time measure. This issue really involves the interaction of the procedure construction and application components of the solution process. A better understanding of what improves with practice might allow a separation of the effects that are due to procedure construction from the effects that result from a more global improvement (e.g., familiarity with the task). A third factor that may complicate the practice effect analysis is that subjects may find different types of procedures more difficult to apply to the same task, differentially affecting the practice effects across conditions. For instance, example-generated procedures might be more difficult to execute since they are relatively new to the subject, and so execution might never be comparable to instruction-procedure execution. (Sternberg and Ketron (1982) reported a similar difficulty in a study of the solution of analogy problems.) In sum, the Computer test and the way it was presented may not be optimal for testing the practice effect hypothesis.

The third hypothesis was more concerned with the pattern of information use. There was some evidence that example subjects do more looking back, especially in the initial stages of the test, which supports the idea that a comparison process is being used. It was also predicted that example subjects would spend less time on the example initially, but refer back more often than instruction subjects. Although the example subjects did look back more often and at different points than instruction subjects, the initial and total amounts of time spent on these sources of information were similar. However, the way that time was used may have been

quite different: when initial time was calculated as amount of time spent on each word, it was evident that the example subjects were spending more time per word. Furthermore, Schematic Instruction time was related to reading speed, and specific example time was not. In conclusion, there was considerable evidence that the patterns of instruction and example information use are different. Two assumptions of the model were supported. There was evidence that retrieval of a procedure was related to performance in the Instruction conditions, since performance suffered when a label was not provided for a procedure. Second, the pattern of references back to the instructions or example supported the idea that examples are compared to the problems, and the time differences suggested that examples were involved in some process that was different than that used for instructions.

The fourth hypothesis was that instruction-generated procedures would be more general than example-generated procedures, that is, they would include more information from a higher-level of the procedure hierarchy. As expected, the instruction protocols were rated as more "general" overall, and they contained fewer specific examples than the example protocols. Furthermore, example subjects seemed less likely to verbalize an abstract procedure and they were more tied to specific problem characteristics. However, the logical extension of this hypothesis, namely that instruction procedures should be more readily transfered to a similar task, was not supported. In fact, the opposite seemed to be true. Example subjects were more likely to transfer a consistent series procedure from the computer task to the series-classification figures task. What does this result imply about the model? Instruction subjects may have retrieved and instantiated a procedure but they could have forgotten the high-level information during the performance of the test, retaining only the necessary specific information in their procedures. At the end of the test, they would have a low-level procedure that may be difficult to apply to the Transfer test because it is in fact too task-specific.

Anecdotal evidence supports this conjecture. One instruction subject who was completely consistent on the Computer test but did not transfer his procedure said that he had "forgotten the instructions" by the time he was asked to do the

transfer test. The fact that instruction subjects' protocols were quite general may actually reflect residual memory for the words in the instructions. The protocols might have little to do with the procedure that was used. This might also explain why there was little difference in the quality of the protocols given by instruction users and non-users, since the protocol may not have reflected the procedure that subjects were actually using.

Conversely, although the procedure constructed by example subjects may also be task-specific and difficult to apply to the Transfer test, they had an advantage in that they may have benefitted from the fact that the Transfer test was quite similar to the first test. Example subjects seemed to retain a greater amount of specific problem information, and they may have had their analogical mechanism primed by experience with the previous example. They may have been able to use this mechanism in conjunction with the new task problems and a remembered problem to start the construction of a new procedure. Therefore, the original generality of the procedure used on the computer task may be almost irrelevant for Transfer performance if the important difference between example and instruction subjects result from differences in the processes they used to construct their procedures. This counter-intuitive finding for Transfer performance is worth pursing. It may be an important key to the difference between example and instruction procedures.

Overall, the results of this study supported the model of instruction and example use described in the introduction. First, there was support for the idea that instructions are used to retrieve procedures. If the appropriate label or level of action is not specified, subjects find it difficult to retrieve an appropriate procedure. Furthermore, there was some evidence that instruction-procedures were more general or schematic than example-procedures, although this did not result in any visible performance advantages for this task.

There was also support for the idea that examples are used in conjunction with an analogical mechanism. First, the similarity of an item to an example was an important determinant of accurate performance, as predicted. Second, example

subjects also looked back more often, and earlier in the course of solving the problems than instruction subjects, presumably to assist procedure construction. Third, the time spent on an example reflects something other than its length or the subject's reading speed, such as procedure construction. Fourth, successful example use was related to inductive ability (of which analogical reasoning is one type). Thus, it may be concluded that subjects do use examples comparatively and analogically, and this use is reflected in subsequent performance. The proposed model adequately accounts for most of the observed performance differences.

An alternative explanation that can also account for at least some of the observed differences between instructions and examples is that the example subjects may have simply taken longer than the instructions subjects to retrieve an appropriate procedure. Thus, the practice effect would represent this "search" for the correct procedure. This account is consistent with the fact that the subjects in the Procedural Instruction condition had some difficulty retrieving procedure, and they also showed a practice effect. However, this explanation does not satisfactorily account for the differences in Inductive-Reasoning performance that were evident for example users and non-users. If retrieval simply takes longer, then it is not clear why the poor inducers do not retrieve a procedure in the Example conditions, although they may be able to do so in the Instruction conditions. Perhaps poor inducers do well with Ordinal examples because the obvious nature of the "series" allows them to retrieve a procedure and bypass construction. The similarity effect also suggests that the example has some direct influence on the subject's procedure throughout the course of the test. The crucial test of this alternate explanation requires a task in which performance is measurably different for a retrieved vs a constructed procedure (cf., Rumelhart & Norman, 1981). In this vein, the Transfer test differences suggest that the procedures were different, although it did not pinpoint the exact nature of the difference.

There are at least two areas to which it may be important to extend the model. The first concerns the interaction between procedure construction and procedure application. Since instruction and example procedures involve differing

types and quantities of constructive activity, it may be important to know the way in which construction and application influence performance when both are occurring simultaneously. One way of characterizing the instruction-example performance differences is to say that application supersedes construction sooner for instruction than for example subjects. For instance, it is likely that while application was operating near the end of the test for all subjects, at the beginning of the test, example subjects may have been doing mostly construction, and instruction subjects some combination of instantiation (which may be like minimal construction) and application. If this is true, it may have implications for the number of practice trials that should be given as a function of the type of instructional information. Although this may be an empirically determinable variable, there is evidence from the schema abstraction literature that the presentation order of information may affect procedure construction for example subjects (Elio & Anderson, 1984). An example procedure may only be as successful as the variety of problems that are used to construct it.

The second important area to which the model should be extended is the situation in which subjects are given both instructions and an example. Since the use of these information sources is quite different, the way in which they are used interactively may be quite important for performance. Previous research has demonstrated that people tend to ignore instructions in favour of examples, in part because they believe that examples are better and easier to understand (LeFevre & Dixon, 1984). There was no evidence that this was objectively true for the instruction and example information used in this study, since overall, both were moderately useful. However, if this belief in an example advantage is widespread, then it may be important to emphasize the instructions or to mesh the instructions and the examples together in such a way that they complement each other, since both are important sources of information. The present research has shown that instructions and example provide different aspects of a procedure. If the instructions can be used to retrieve a known procedure, and the example used to fill in the specific, task-relevant actions, then perhaps optimal performance will result.

In the Introduction, it was suggested that there are at least three important ways in which instruction sets can vary: the amount of information provided, whether or not a procedure was described, and how representative the information was of the task. Of these, only the third was demonstrated to have any appreciable effect on performance in this study. For instruction, representativeness is a matter of providing a procedure at the right hierarchical level for the particular task, and for examples, representativeness is the "best" type of example. This is an important demonstration, and it supports the hypothesis that procedure retrieval or construction is important for performance. There are a number of other issues that were raised by the study that deserve further research. One is the suggestion that inductive ability might be related to example use. This deserves further study because it indicates that subject factors may influence procedure construction, as well as application. Some of the relevant subject factors might include reading ability, inductive ability, experience with similar tasks, previous beliefs about instruction and example information, and prior knowledge. Second, it may also be important to study task variables that could affect performance. The most obvious one is complexity, but the purpose of the task for the subject may also be important. Third, the type of performance that is required will be another important influence on the instruction and example information, and should be taken into account by a complete model of instruction/example use.

Bibliography

Anderson, J. R., Farrell, R. & Sauers, R. (1983). Learning to program in LISP. (ONR Technical Report, ONR-83-1). Pittsburgh: Carnegie-Mellon University.

Anderson, J. R., Farrell, R. & Sauers, R. (1984). Learning to program in LISP. Cognitive Science, 8, 87-129.

Anderson, J. R., Greeno, J. G., Kline, P. J. & Neves, D. M. (1981). Acquisition of problem-solving skill. In J. R. Anderson (Ed.) Cognitive Skills and their Acquisition. (pp. 191-230). Hillsdale, N. J.: Erlbaum.

Barnard, P. S., Wright, P., & Wilcox, P. (1979). Effects of response instructions and question style on the ease of completing forms. Journal of Occupational Psychology, 52 209-226.

Bisanz, J., Bisanz, G. L., & LeFevre, J., (1984). Interpretation of instructions: A source of individual differences in analogical reasoning. Intelligence, 8, 161-177.

Bransford, J. D. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-721.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.) Cognition and Categorization (pp. 169-211). Hillsdale, N. J.: Erlbaum.

Dixon, P. (1982). Plans and written directions for complex tasks. Journal of Verbal Learning and Verbal Behavior, 21, 70-84.

Dixon, P. (in press). Actions and Procedural Directions. In R. Tomlin (Ed.) Coherence and grounding in discourse. Amsterdam: John Benjamins.

Elio, R. & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. Journal of Experimental Psychology: Human Learning and Memory, 7, 397-417.

Elio, R. & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. Memory and Cognition, 12, 20-30.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1 - 38.

Hayes-Roth, F. (1978). Learning by example. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.) Cognitive psychology and instruction (pp. 27-38). New York: Plenum Press.

Hayes-Roth, B. & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-338

Jones, S. (1966). The effect of a negative qualifier in an instruction. Journal of Verbal Learning and Verbal Behavior, 5, 497-501.

Kenny, D. A. & Smith, E. R. (1980). A note on the analysis of designs in which subjects receive each stimulus only once. Journal of Experimental Social Psychology, 16, 497-507.

LeFevre, J. & Dixon, P. (1984). Do written instructions need examples? Paper submitted for review.

Lorge, I. & Thorndike, R. L. (1954). The Lorge-Thorndike Intelligence Test, Levels A-H. Boston: Houghton-Mifflin Company.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 333-352.

Medin, D. L. & Schaffer, M. M. (1978). A context theory of classification learning. Psychological Review, 85, 207-238.

Resnick, L. B. & Glaser, R. (1976). Problem solving and intelligence. In L. B. Resnick (Ed.) The nature of intelligence (pp. 205-230). Hillsdale, N. J.: Lawrence Erlbaum Associates.

Rumelhart, D. E. & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.) Cognitive skills and their acquisition. (pp. 335-360). Hillsdale, N. J.: Erlbaum.

Simon, H. A. & Hayes, J. R. (1976). Understanding complex instructions. In D. Klahr (Ed.) Cognition and instruction (pp. 269-285). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Smith, E. E. & Goodman, L. (1982). Understanding instructions: The role of explanatory material. (Report No. 5088, Bolt, Beranek and Newman Inc.)

Sternberg, R. J. (1977). Intelligence, information-processing and analogical reasoning: The componential analysis of human abilities. Hillsdale, N. J.: Erlbaum.

Sternberg, R. J. (1981). Intelligence and non-entrenchment. Journal of Educational Psychology, 73, 1-16.

Sternberg, R. J. & Ketron, J. L. (1982). Selection and implementation of strategies in reasoning by analogy. Journal of Educational Psychology, 74, 399-413.

Sternberg, R. J. & Gardner, M. K. (1983). Unities in inductive reasoning. Journal of Experimental Psychology: General, 112, 80-116.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tourangeau, R. & Sternberg, R. J. (1982). Understanding and appreciating metaphors. Cognition, 11, 203-244.

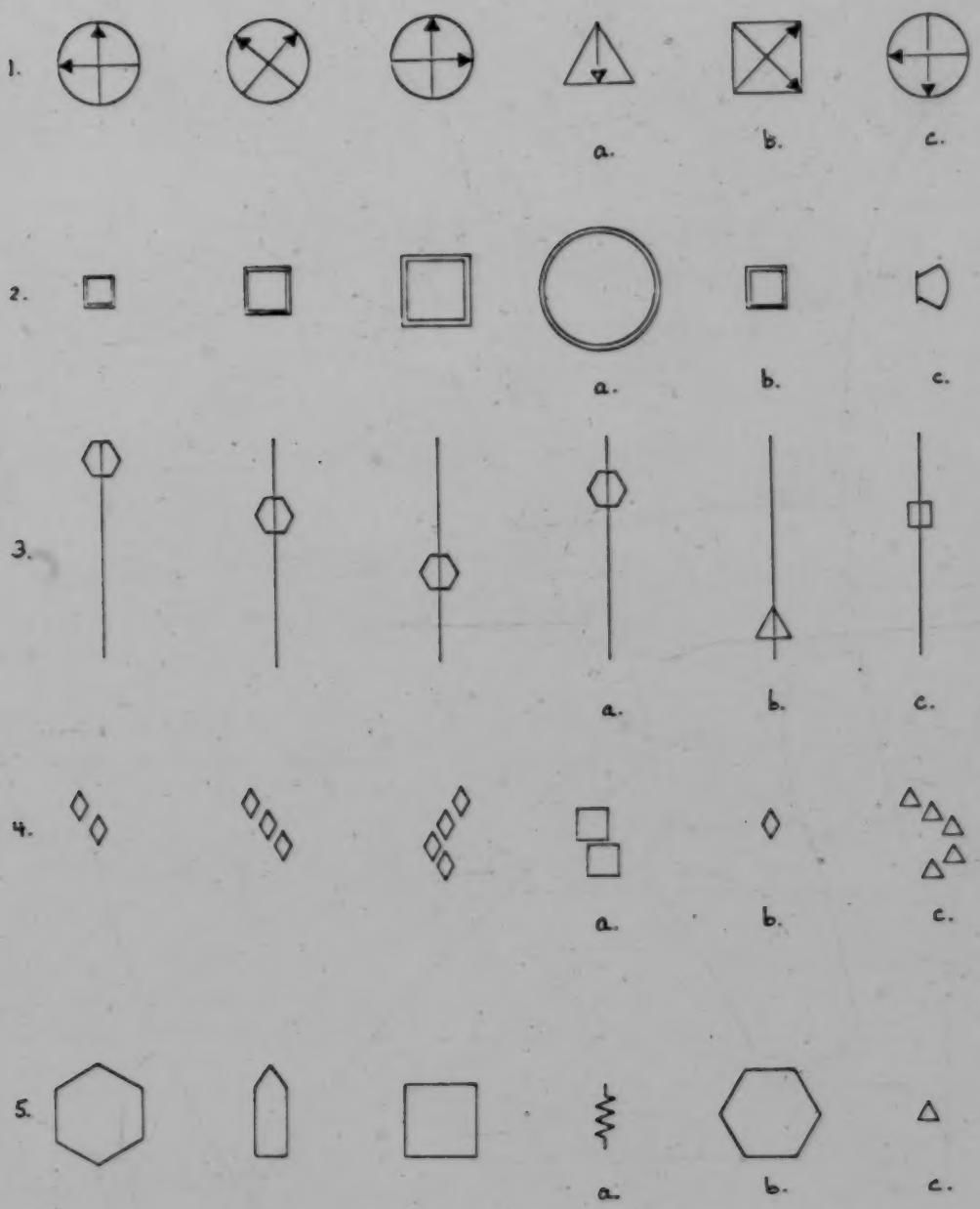
Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill Book Company.

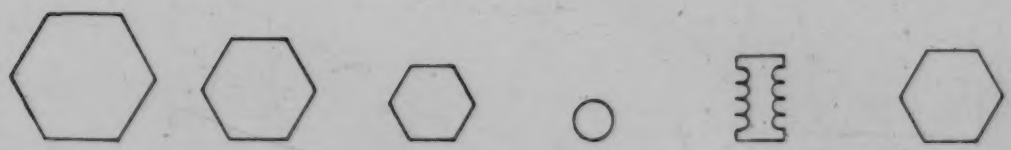
Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. Instructional Science, 7, 249-312.

Wright, P. (1981). "The instructions clearly state..." Can't people read? Applied Ergonomics, 12, 131-141.

APPENDIX 1

Series-Classification Test

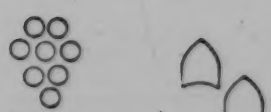

1.
a.
b.
c.
2.
a.
b.
c.
3.
a.
b.
c.
4.
a.
b.
c.
5.
a.
b.
c.

6. 

a. b. c.

7.

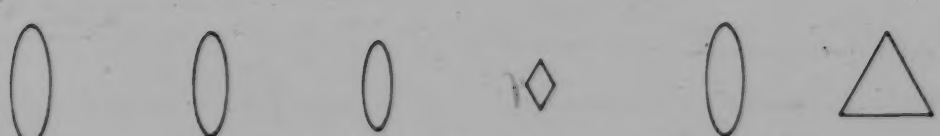
a. b. c.

8. 2 4 6  3

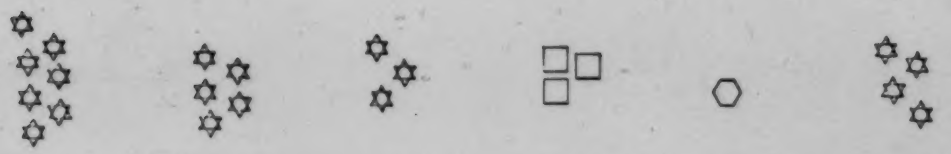
a. b. c.

9.

a. b. c.

10. 

a. b. c.

11. 

a. b. c.

12.

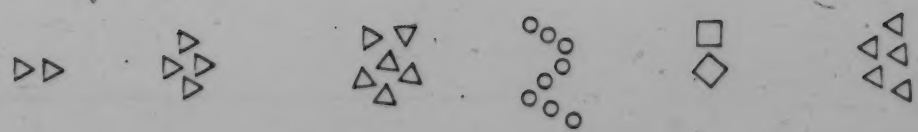
a. b. c.

13.

a. b. c.

14.

a. b. c.

15. 

a. b. c.

APPENDIX 2

Supplementary One-Way Analyses

SOURCE	SS	Df	MS	F	PROB.
Initial Time Effect					
Condition	9.1874E8	3	3.0625E8	1.21	.3132
Error	1.7227E10	68	2.5334E8		
Total Time Effect					
Condition	4.3757E11	3	1.4586E11	2.42	.0733
Error	4.0940E10	68	6.0207E8		
Number of References					
Condition	22.153	3	7.3843	2.41	.0749
Error	208.722	68	3.0694		
Accuracy					
Condition	9.9235	3	3.3078	0.90	.4467
Error	175.8457	48	3.6634		
Problems	127.3306	3	42.4435	28.29	.0000
P by C	7.2438	9	0.8048	0.54	.8461
Error	216.0638	144	1.5044		
Solution Time					
Condition	5.9737E7	3	1.9912E7	0.57	.6400
Error	1.6882E9	48	3.5170E7		
Problems	2.6921E8	3	8.9738E7	9.99	.0000
P by C	7.2990E7	9	8.1100E6	0.90	.5246
Error	1.2934E9	144	8.9821E6		

SOURCE	SS	Df	MS	F	PROB.
Practice					
Condition	3.2855E8	3	1.0951E8	1.25	.3035
Error	4.2191E9	48	8.7898E7		
Practice	5.1889E8	1	5.1889E8	18.33	.0001
Prac by C	1.3878E8	3	4.6262E7	1.63	.1938
Error	1.3586E9	48	2.8304E7		
Problems	5.4282E8	3	1.8128E8	9.31	.0000
Prob by C	7.3134E7	9	8.1260E6	0.42	.9242
Prac by Prob	8.5223E6	3	2.8409E6	0.14	.9362
Prac by Prob by C	1.6701E8	9	1.8557E8	0.91	0.5168
Error	2.9306E9	144	2.0351E7		